Stability of Noisy Metropolis–Hastings

F. J. Medina-Aguayo, A. Lee, and G. O. Roberts

University of Warwick

March 25, 2015

Abstract

Pseudo-marginal Markov chain Monte Carlo methods for sampling from intractable distributions have gained recent interest and have been theoretically studied in considerable depth. Their main appeal is that they are exact, in the sense that they target marginally the correct invariant distribution. However, the pseudo-marginal Markov chain can exhibit poor mixing and slow convergence towards its target. As an alternative, a subtly different Markov chain can be simulated, where better mixing is possible but the exactness property is sacrificed. This is the noisy algorithm, initially conceptualised as Monte Carlo within Metropolis (MCWM), which has also been studied but to a lesser extent. The present article provides a further characterisation of the noisy algorithm, with a focus on fundamental stability properties like positive recurrence and geometric ergodicity. Sufficient conditions for inheriting geometric ergodicity from a standard Metropolis–Hastings chain are given, as well as convergence of the invariant distribution towards the true target distribution.

Keywords: Markov chain Monte Carlo; Pseudo-marginal Monte Carlo; Monte Carlo within Metropolis; Intractable likelihoods; Geometric ergodicity.

1 Introduction

1.1 Intractable target densities and the pseudo-marginal algorithm

Suppose our aim is to simulate from an intractable probability distribution π for some random variable X, which takes values in a measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. In addition, let π have a density $\pi(x)$ with respect to some reference measure $\mu(dx)$, e.g. the counting or the Lebesgue measure. By intractable we mean that an analytical expression for the density $\pi(x)$ is not available and so implementation of a Markov chain Monte Carlo (MCMC) method targeting π is not straightforward.

One possible solution to this problem is to target a different distribution on the extended space $(\mathcal{X} \times \mathcal{W}, \mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{W}))$, which admits π as marginal distribution. The pseudo-marginal algorithm (Beaumont 2003, Andrieu and Roberts 2009) falls into this category since it is a Metropolis–Hastings (MH) algorithm targeting a distribution $\bar{\pi}_N$, associated to the random vector (X, W) defined on the product space $(\mathcal{X} \times \mathcal{W}, \mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{W}))$ where $\mathcal{W} \subseteq \mathbb{R}_0^+ := [0, \infty)$. It is given by

$$\bar{\pi}_N(dx, dw) := \pi(dx) Q_{x,N}(dw) w, \tag{1}$$

where $\{Q_{x,N}\}_{(x,N) \in \mathcal{X} \times \mathbb{N}^+}$ is a family of probability distributions on $(\mathcal{W}, \mathcal{B}(\mathcal{W}))$ satisfying for each $(x, N) \in \mathcal{X} \times \mathbb{N}$

$$\mathbb{E}_{Q_{x,N}}[W_{x,N}] \equiv 1, \quad \text{for } W_{x,N} \sim Q_{x,N}(\cdot). \tag{2}$$

Throughout this article, we restrict our attention to the case where for each $x \in \mathcal{X}$, $W_{x,N}$ is $Q_{x,N}$-a.s. strictly positive, for reasons that will become clear.

The random variables $\{W_{x,N}\}_{x,N}$ are commonly referred as the weights. Formalising this algorithm using (1) and (2) was introduced by Andrieu and Vihola (2015), and "exactness" follows immediately: $\bar{\pi}$ admits π as a marginal. Given a proposal kernel $q : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0, 1]$, the respective proposal of the pseudo-marginal is given by

$$\bar{q}_N(x, w; dy, du) := q(x, dy) Q_{y,N}(du),$$

and, consequently, the acceptance probability can be expressed as

$$\bar{\alpha}_N(x, w; y, u) := \min\left\{1, \frac{\pi(dy)uq(y, dx)}{\pi(dx)wq(x, dy)}\right\}. \tag{3}$$

The pseudo-marginal algorithm defines a time-homogeneous Markov chain, with transition kernel $\bar{P}_N$ on the measurable space $(\mathcal{X} \times \mathcal{W}, \mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{W}))$. A single draw from $\bar{P}_N(x, w; \cdot, \cdot)$ is presented in Algorithm 1.

Algorithm 1 Simulating from $\bar{P}_N(x, w; \cdot, \cdot)$

---

1. Sample $Y \sim q(x, \cdot)$.

2. Draw $U \sim Q_{Y,N}(\cdot)$.

3. With probability $\bar{\alpha}_N(x, w; Y, U)$ defined in (3):
   return (Y, U),
   otherwise:
   return (x, w).

Due to its exactness and straightforward implementation in many settings, the pseudo-marginal has gained recent interest and has been theoretically studied in some depth, see e.g. Andrieu and Roberts (2009), Andrieu and Vihola (2015), Andrieu and Vihola (2014), Sherlock et al. (2015), Girolami et al. (2013) and Maire et al. (2014). These studies typically compare the pseudo-marginal Markov chain with a "marginal" Markov chain, arising in the case where all the weights are almost surely equal to 1, and (3) is then the standard Metropolis–Hastings acceptance probability associated with the target density π and the proposal q.

1.2 Examples of pseudo-marginal algorithms

A common source of intractability for π occurs when a latent variable Z on $(Z, \mathcal{B}(Z))$ is used to model observed data, as in hidden Markov models (HMMs) or mixture models. Although the density $\pi(x)$ cannot be computed, it can be approximated via importance sampling, using an appropriate auxiliary distribution, say ν_x. Here, appropriate means $\pi_x \ll \nu_x$, where π_x denotes the conditional distribution of Z given $X = x$. Therefore, for this setting, the weights are given by

$$W_{x,N} = \frac{1}{N} \sum_{k=1}^{N} \frac{\pi_x\left(Z_x^{(k)}\right)}{\nu_x\left(Z_x^{(k)}\right)}, \quad \text{where } \left\{Z_x^{(k)}\right\}_{k \in \{1,\ldots,N\}} \overset{i.i.d.}{\sim} \nu_x(\cdot),$$

which motivates the following generic form when using averages of unbiased estimators

$$W_{x,N} = \frac{1}{N} \sum_{k=1}^{N} W_x^{(k)}, \quad \text{where } \left\{W_x^{(k)}\right\}_k \overset{i.i.d.}{\sim} Q_x(\cdot), \mathbb{E}_{Q_x}\left[W_x^{(k)}\right] \equiv 1. \tag{4}$$

It is clear that (4) describes only a special case of (2). Nevertheless, we will pay special attention to the former throughout the article. For similar settings to (4) see Andrieu and Roberts (2009).

Since (2) is more general, it allows $W_{x,N}$ to be any random variable with expectation 1. Sequential Monte Carlo (SMC) methods involve the simulation of a system of some number of particles, and provide unbiased estimates of likelihoods associated with HMMs (see Del Moral 2004, Proposition 7.4.1 or Pitt et al. 2012) irrespective of the size of the particle system. Consider the model given by Figure 1. The
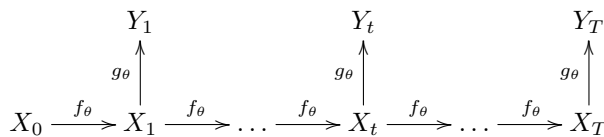

Figure 1: Hidden Markov Model.

random variables $\{X_t\}_{t=0}^{T}$ form a time-homogeneous Markov chain with transition $f_\theta(\cdot|x_{t-1})$ that depends

on a set of parameters θ. The observed random variables $\{Y_t\}_{t=1}^T$ are conditionally independent given the unobserved $\{X_t\}_{t=1}^T$ and are distributed according to $g_\theta(\cdot|x_t)$, which also may depend on θ. The likelihood function for θ is given by

$$l(\theta; y_1, \ldots, y_T) := \mathbb{E}_{f_\theta}\left[\prod_{t=1}^T g_\theta(y_t|X_t)\right],$$

where $\mathbb{E}_\theta$ denotes expectation w.r.t. the θ-dependent law of $\{X_t\}_{t=1}^T$, and we assume for simplicity that the initial value $X_0 = x_0$ is known. If we denote by $\hat{l}_N(\theta; y_1, \ldots, y_T)$ the unbiased SMC estimator of $l(\theta; y_1, \ldots, y_T)$ based on N particles, we can then define

$$W_{\theta, N} := \frac{\hat{l}_N(\theta; y_1, \ldots, y_T)}{l(\theta; y_1, \ldots, y_T)},$$

and (2) is satisfied but (4) is not. The resulting pseudo-marginal algorithm is developed and discussed in detail in Andrieu et al. (2010), where it and related algorithms are referred to as particle MCMC methods.

1.3   The noisy algorithm

Although the pseudo-marginal has the desirable property of exactness, it can suffer from "sticky" behaviour, exhibiting poor mixing and slow convergence towards the target distribution (Andrieu and Roberts 2009 and Lee and Łatuszyński 2014). The cause for this is well-known to be related with the value of the ratio between $W_{y,N}$ and $W_{x,N}$ at a particular iteration. Heuristically, when the value of the current weight (w in (3)) is large, proposed moves can have a low probability of acceptance. As a consequence, the resulting chain can get "stuck" and may not move after a considerable number of iterations.

In order to overcome this issue, a subtly different algorithm is performed in some practical problems (see, e.g., McKinley et al., 2014). The basic idea is to refresh, independently from the past, the value of the current weight at every iteration. The ratio of the weights between $W_{y,N}$ and $W_{x,N}$ still plays an important role in this alternative algorithm, but here refreshing $W_{x,N}$ at every iteration can improve mixing and the rate of convergence.

This alternative algorithm is commonly known as Monte Carlo within Metropolis (MCWM), as in O'Neill et al. (2000), Beaumont (2003) or Andrieu and Roberts (2009), since typically the weights are Monte Carlo estimates as in (4). From this point onwards it will be referred as the noisy MH algorithm or simply the noisy algorithm to emphasize that our main assumption is (2). Due to independence from previous iterations while sampling $W_{x,N}$ and $W_{y,N}$, the noisy algorithm also defines a time-homogeneous Markov chain with transition kernel $\tilde{P}_N$, but on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. A single draw from $\tilde{P}_N(x, \cdot)$ is presented in Algorithm 2, and it is clear that we restrict our attention to strictly positive weights because the algorithm is not well-defined when both $W_{y,N}$ and $W_{x,N}$ are equal to 0.

Algorithm 2 Simulating from $\tilde{P}_N(x, \cdot)$

1. Sample $Y \sim q(x, \cdot)$.

2. Draw $W \sim Q_{x,N}(\cdot)$ and $U \sim Q_{Y,N}(\cdot)$, independently.

3. With probability $\bar{\alpha}_N(x, W; Y, U)$ defined in (3):
        return Y,
     otherwise:
        return x.

Even though these algorithms differ only slightly, the related chains have very different properties. In Algorithm 2, the value w is generated at every iteration whereas in Algorithm 1, it is treated as an input. As a consequence, Algorithm 1 produces a chain on $(\mathcal{X} \times \mathcal{W}, \mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{W}))$ contrasting with a chain from Algorithm 2 taking values on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. However, the noisy chain is not invariant under π and it is not reversible in general. Moreover, it may not even have an invariant distribution as shown by some examples in Section 2.
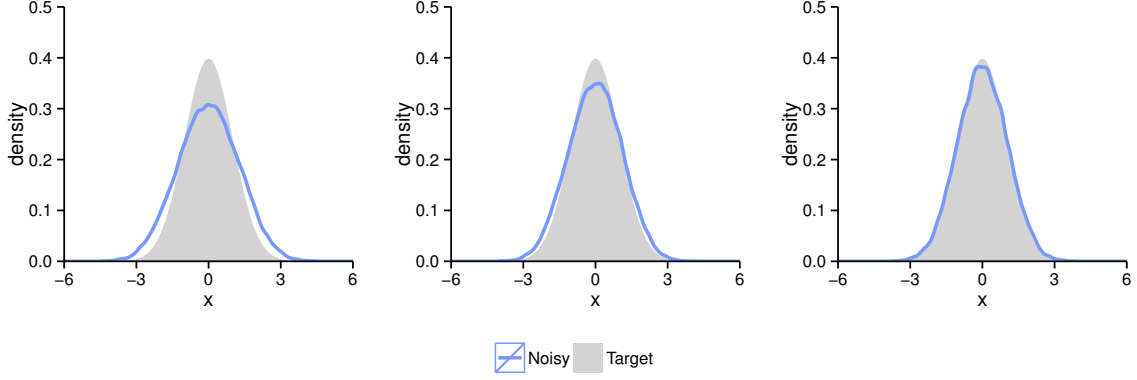
Figure 2: Estimated densities using the noisy chain with $100,000$ iterations for $N = 10$ (left), $N = 100$ (central) and $N = 1,000$ (right).

From O'Neill et al. (2000) and Fernández-Villaverde and Rubio-Ramírez (2007), it is evident that the implementation of the noisy algorithm goes back even before the appearance of the pseudo-marginal, the latter initially conceptualised as Grouped Independence Metropolis–Hastings (GIMH) in Beaumont (2003). Theoretical properties, however, of the noisy algorithm have mainly been studied in tandem with the pseudo-marginal by Beaumont (2003), Andrieu and Roberts (2009) and more recently by Alquier et al. (2014).

1.4 Objectives of the article

The objectives of this article can be illustrated using a simple example. Let $\mathcal{N}(\cdot|\mu, \sigma^2)$ denote a univariate Gaussian distribution with mean μ and variance σ^2 and $\pi(\cdot) = \mathcal{N}(\cdot|0, 1)$ be a standard normal distribution. Let the weights $W_{x,N}$ be as in (4) with

$$Q_x(\cdot) = \log \mathcal{N}\left(\cdot \left| -\frac{1}{2}\sigma^2, \sigma^2\right.\right) \quad \text{and} \quad \sigma^2 := 5,$$

where $\log \mathcal{N}(\cdot|\mu, \sigma^2)$ denotes a log-normal distribution of parameters μ and σ^2. In addition, let the proposal q be random walk given by $q(x, \cdot) = \mathcal{N}(\cdot|x, 4)$. For this example, Figure 2 shows the estimated densities using the noisy chain for different values of N. It appears that the noisy chain has an invariant distribution, and as N increases it seems to approach the desired target π. Our objectives here are to answer the following types of questions about the noisy algorithm in general:

1. Does an invariant distribution exist, at least for N large enough?

2. Does the noisy Markov chain behave like the marginal chain for sufficiently large N?

3. Does the invariant distribution, if it exists, converge to π as N increases?

We will see that the answer to the first two questions is negative in general. However, all three questions can be answered positively when the marginal chain is geometrically ergodic and the distributions of the weights satisfy additional assumptions.

1.5 Marginal chains and geometric ergodicity

In order to formalise our analysis, let P denote the Markov transition kernel of a standard MH chain on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, targeting π with proposal q. We will refer to this chain and this algorithm using the term marginal (as in Andrieu and Roberts 2009 and Andrieu and Vihola 2015), which can be seen as an idealised version for which the noisy chain and corresponding algorithm are simple approximations. Therefore

$$P(x, dy) := \alpha(x, y)q(x, dy) + \delta_x(dy)\rho(x),$$

4

where α is the MH acceptance probability and ρ is the rejection probability, given by

$$\alpha(x, y) := \min\left\{1, \frac{\pi(dy)q(y, dx)}{\pi(dx)q(x, dy)}\right\} \quad \text{and} \quad \rho(x) := 1 - \int_{\mathcal{X}} \alpha(x, y)q(x, dy). \tag{5}$$

Similarly, for the transition kernel $\tilde{P}_N$ of the noisy chain, moves are proposed according to q but are accepted using $\bar{\alpha}_N$ (as in (3)) instead of α, once values for $W_{x,N}$ and $W_{y,N}$ are sampled. In order to distinguish the acceptance probabilities between the noisy and the pseudo-marginal processes, despite being the same after sampling values for the weights, define

$$\tilde{\alpha}_N(x, y) := \mathbb{E}_{Q_{x,N} \otimes Q_{y,N}} \bar{\alpha}_N(x, W_{x,N}; y, W_{y,N}). \tag{6}$$

Here $\tilde{\alpha}_N$ is the expectation of a randomised acceptance probability, which permits defining the transition kernel of the noisy chain by

$$\tilde{P}_N(x, dy) := \tilde{\alpha}_N(x, y)q(x, dy) + \delta_x(dy)\tilde{\rho}_N(x),$$

where $\tilde{\rho}_N$ is the noisy rejection probability given by

$$\tilde{\rho}_N(x) := 1 - \int_{\mathcal{X}} \tilde{\alpha}_N(x, y)q(x, dy). \tag{7}$$

The noisy kernel $\tilde{P}_N$ is just a perturbed version of P, involving a ratio of weights in the noisy acceptance probability $\tilde{\alpha}_N$. In addition, when such weights are identically one, i.e. $Q_{x,N}(\{1\}) = 1$, the noisy chain reduces to the marginal chain, whereas the pseudo-marginal becomes the marginal chain with an extra component always equal to 1.

So far, the terms slow convergence and "sticky" behaviour have been used in a relative vague sense. A powerful characterisation of the behaviour of a Markov chain is provided by geometric ergodicity, defined below. Geometrically ergodic Markov chains have a limiting invariant probability distribution, which they converge towards geometrically fast in total variation (Meyn and Tweedie, 2009). For any Markov kernel $K : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0, 1]$, let K^n be the n-step transition kernel, which is given by

$$K^n(x, \cdot) := \int_{\mathcal{X}} K^{n-1}(x, dz)K(z, \cdot), \quad \text{for } n \geq 2.$$

Definition 1.1 (Geometric ergodicity). A φ-irreducible and aperiodic Markov chain $\boldsymbol{\Phi} := (\Phi_i)_{i \geq 0}$ on a measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, with transition kernel P and invariant distribution π, is geometrically ergodic if there exists a finite function $V \geq 1$ and constants $\tau < 1$, $R < \infty$ such that

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{TV} \leq RV(x)\tau^n, \quad \text{for } x \in \mathcal{X}. \tag{8}$$

Here, $\| \cdot \|_{TV}$ denotes the total variation norm given by

$$\|\mu\|_{TV} = \frac{1}{2} \sup_{|g| \leq 1} \left| \int \mu(dy)g(y) \right| = \sup_{A \in \mathcal{B}(\mathcal{X})} \mu(A),$$

where μ is any signed measure.

Geometric ergodicity does not necessarily provide fast convergence in an absolute sense. For instance, consider cases where τ, or R, from Definition 1.1 are extremely close to one, or very large respectively. Then the decay of the total variation distance, though geometric, is not particularly fast (see Roberts and Rosenthal 2004 for some examples).

Nevertheless, geometric ergodicity is a useful tool when analysing non-reversible Markov chains as will become apparent in the noisy chain case. Moreover, in practice one is often interested in estimating $\mathbb{E}_\pi[f(X)]$ for some function $f : \mathcal{X} \to \mathbb{R}$, which is done by using ergodic averages of the form

$$e_{n,m}(f) := \frac{1}{n} \sum_{i=m+1}^{m+n} f(\Phi_i), \quad \text{for } m, n \geq 0.$$

In this case, geometric ergodicity is a desirable property since it can guarantee the existence of a central limit theorem (CLT) for $e_{n,m}(f)$, see Chan and Geyer (1994), Roberts and Rosenthal (1997) and Roberts

and Rosenthal (2004) for a more general review. Also, its importance is related with the construction of consistent estimators of the corresponding asymptotic variance in the CLT, as in Flegal and Jones (2010).

As noted in Andrieu and Roberts (2009), if the weights $\{W_{x,N}\}_{x,N}$ are not essentially bounded then the pseudo-marginal chain cannot be geometrically ergodic; in such cases the "stickiness" may be more evident. In addition, under mild assumptions (in particular, that $\bar{P}_N$ has a left spectral gap), from Andrieu and Vihola (2015, Proposition 10) and Lee and Łatuszyński (2014), a sufficient but not necessary condition ensuring the pseudo-marginal inherits geometric ergodicity from the marginal, is that the weights are uniformly bounded. This certainly imposes a tight restriction in many practical problems.

The analyses in Andrieu and Roberts (2009) and Alquier et al. (2014) mainly study the noisy algorithm in the case where the marginal Markov chain is uniformly ergodic, i.e. when it satisfies (8) with $\sup_{x\in\mathcal{X}} V(x) < \infty$. However, there are many Metropolis–Hastings Markov chains for statistical estimation that cannot be uniformly ergodic, e.g. random walk Metropolis chains when π is not compactly supported. Our focus is therefore on inheritance of geometric ergodicity by the noisy chain, complementing existing results for the pseudo-marginal chain.

1.6 Outline of the paper

In Section 2, some simple examples are presented for which the noisy chain is positive recurrent, so it has an invariant probability distribution. This is perhaps the weakest stability property that one would expect a Monte Carlo Markov chain to have. However, other fairly surprising examples are presented for which the noisy Markov chain is transient even though the marginal and pseudo-marginal chains are geometrically ergodic. Section 3 is dedicated to inheritance of geometric ergodicity from the marginal chain, where two different sets of sufficient conditions are given and are further analysed in the context of arithmetic averages given by (4). Once geometric ergodicity is attained, it guarantees the existence of an invariant distribution $\tilde{\pi}_N$ for the noisy chain. Under the same sets of conditions, we show in Section 4 that $\tilde{\pi}_N$ and π can be made arbitrarily close in total variation as N increases. Moreover, explicit rates of convergence are possible to obtain in principle, when the weights arise from an arithmetic average setting as in (4).

2 Motivating examples

2.1 Homogeneous weights with a random walk proposal

Assume a log-concave target distribution π on the positive integers, whose density with respect to the counting measure is given by

$$\pi(m) \propto \exp\{-h(m)\}\, \mathbb{1}_{m\in\mathbb{N}^+},$$

where $h : \mathbb{N}^+ \to \mathbb{R}$ is a convex function. In addition, let the proposal distribution be a symmetric random walk on the integers, i.e.

$$q(m, \{m+1\}) = \frac{1}{2} = q(m, \{m-1\}), \quad \text{for } m \in \mathbb{Z}. \tag{9}$$

From Mengersen and Tweedie (1996), it can be seen that the marginal chain is geometrically ergodic.

Now, assume the distribution of the weights $\{W_{m,N}\}_{m,N}$ is homogeneous with respect to the state space, meaning

$$W_{m,N} = W_N \sim Q_N(\cdot), \quad \text{for all } m \in \mathbb{N}^+. \tag{10}$$

In addition, assume $W_N > 0$ Q_N-a.s., then for $m \geq 2$

$$\tilde{P}_N(m, \{m-1\}) = \frac{1}{2}\mathbb{E}_{Q_N\otimes Q_N}\left[\min\left\{1, \frac{\exp\{h(m)\}}{\exp\{h(m-1)\}} \times \frac{W_N^{(1)}}{W_N^{(2)}}\right\}\right] \quad \text{and}$$

$$\tilde{P}_N(m, \{m+1\}) = \frac{1}{2}\mathbb{E}_{Q_N\otimes Q_N}\left[\min\left\{1, \frac{\exp\{h(m)\}}{\exp\{h(m+1)\}} \times \frac{W_N^{(1)}}{W_N^{(2)}}\right\}\right], \quad \text{where } \left\{W_N^{(k)}\right\}_{k\in\{1,2\}} \overset{i.i.d.}{\sim} Q_N(\cdot).$$

For this particular class of weights and using the fact that h is convex, the noisy chain is geometrically ergodic, implying the existence of an invariant probability distribution.
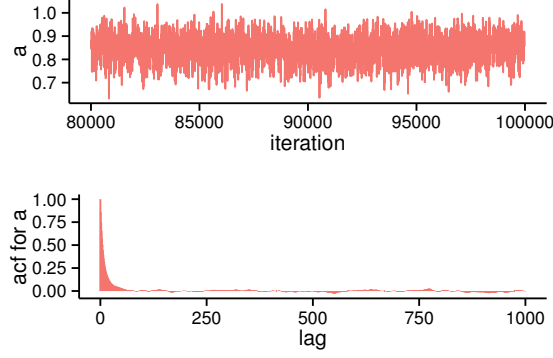
Figure 3: Last $20,000$ iterations of the marginal algorithm for the autoregressive parameter a (top). Estimated autocorrelation function of the corresponding marginal chain (bottom). The mean acceptance probability was 0.256.

Proposition 2.1. *Consider a log-concave target density on the positive integers and a proposal density as in* (9). *In addition, let the distribution of the weights be homogeneous as in* (10). *Then, the chain generated by the noisy kernel $\tilde{P}_N$ is geometrically ergodic.*

It is worth noting that the distribution of the weights, though homogeneous with respect to the state space, can be taken arbitrarily, as long as the weights are positive. Homogeneity ensures that the distribution of the ratio of such weights is not concentrated near 0, due to its symmetry around one, i.e. for $z > 0$

$$\mathbb{P}_{Q_N \otimes Q_N} \left[\frac{W_N^{(1)}}{W_N^{(2)}} \leq z \right] = \mathbb{P}_{Q_N \otimes Q_N} \left[\frac{W_N^{(1)}}{W_N^{(2)}} \geq \frac{1}{z} \right], \quad \text{where} \quad \left\{ W_N^{(k)} \right\}_{k \in \{1,2\}} \overset{i.i.d.}{\sim} Q_N(\cdot).$$

In contrast, when the support of the distribution Q_N is unbounded, the corresponding pseudo-marginal chain cannot be geometrically ergodic.

2.2 Particle MCMC

More complex examples arise when using particle MCMC methods, for which noisy versions can also be performed. They may prove to be useful in some inference problems. Consider again the hidden Markov model given by Figure 1. As before, set $X_0 = x_0$ and let

$$\theta = \left\{ x_0, a, \sigma_X^2, \sigma_Y^2 \right\},$$
$$f_\theta \left(\cdot | X_{t-1} \right) = \mathcal{N} \left(\cdot | a X_{t-1}, \sigma_X^2 \right) \quad \text{and}$$
$$g_\theta \left(\cdot | X_t \right) = \mathcal{N} \left(\cdot | X_t, \sigma_Y^2 \right).$$

Therefore, once a prior distribution for θ is specified, $p(\cdot)$ say, the aim is to conduct Bayesian inference on the posterior distribution

$$\pi(\theta | y_1, \ldots, y_T) \propto p(\theta) l(\theta; y_1, \ldots, y_T).$$

In this particular setting, the posterior distribution is tractable. This will allows us to compare the results obtained from the exact and noisy versions, both relying on the SMC estimator $\hat{l}_N(\theta; y_1, \ldots, y_T)$ of the likelihood. Using a uniform prior for the parameters and a random walk proposal, Figure 3 shows the run and autocorrelation function (acf) for the autoregressive parameter a of the marginal chain. Similarly, Figure 4 shows the corresponding run and acf for both the pseudo-marginal and the noisy chain when $N = 250$. It is noticeable how the pseudo-marginal gets "stuck", resulting in a lower acceptance than the marginal and noisy chains. In addition, the acf of the noisy chain seems to decay faster than that of the pseudo-marginal chain.

Finally, Figure 5 and Figure 6 show the estimated posterior densities for the parameters when $N = 250$ and $N = 750$, respectively. There, the trade-off between the pseudo-marginal and the noisy algorithm is noticeable. For lower values of N, the pseudo-marginal will require more iterations due to the slow mixing, whereas the noisy converges faster towards an unknown noisy invariant distribution. By increasing N, the mixing in the pseudo-marginal improves and the noisy invariant approaches the true posterior.
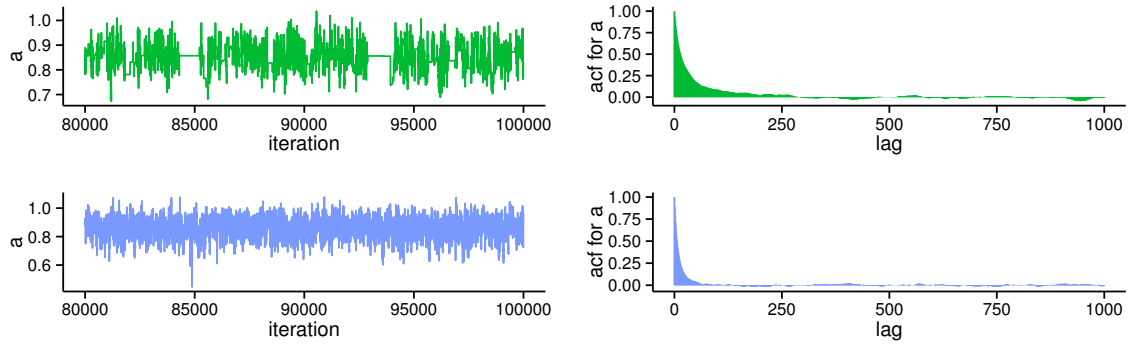
Figure 4: Last 20,000 iterations of the pseudo-marginal (top left) and noisy (bottom left) algorithms, for the autoregressive parameter a when $N = 250$. Estimated autocorrelation functions of the corresponding pseudo-marginal (top right) and noisy (bottom right) chains. The mean acceptance probabilities were 0.104 for the pseudo-marginal and 0.283 for the noisy chain.

Figure 5: Estimated densities using the marginal, pseudo-marginal and noisy chains for the 4 parameters when $N = 250$. Vertical lines indicate the real values.
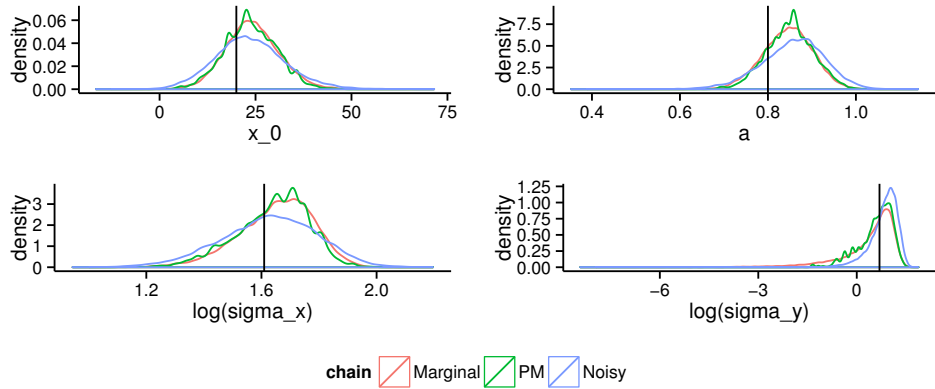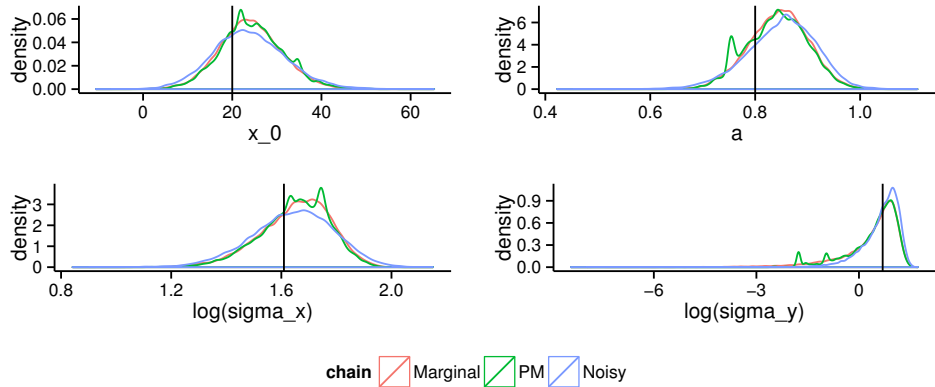

Figure 6: Estimated densities using the marginal, pseudo-marginal and noisy chains for the 4 parameters, when $N = 750$. Vertical lines indicate the real values.

2.3 Transient noisy chain with homogeneous weights

In contrast with example in Section 2.1, this one shows that the noisy algorithm can produce a transient chain even in simple settings. Let π be a geometric distribution on the positive integers, whose density with respect to the counting measure is given by

$$\pi(m) = \left(\frac{1}{2}\right)^m \mathbb{1}_{\{m \in \mathbb{N}^+\}}. \tag{11}$$

In addition, assume the proposal distribution is a simple random walk on the integers, i.e.

$$q(m, \{m+1\}) = \theta = 1 - q(m, \{m-1\}), \quad \text{for } m \in \mathbb{Z}. \tag{12}$$

where $\theta \in (0, 1)$. Under these assumptions, the marginal chain is geometrically ergodic, see Proposition A.1 in Appendix A.

Consider $N = 1$ and as in Section 2.1, let the distribution of weights be homogeneous and given by

$$W = (b - \varepsilon)Ber(s) + \varepsilon, \quad \text{for } b > 1 \text{ and } \varepsilon \in (0, 1), \tag{13}$$

where $Ber(s)$ denotes a Bernoulli random variable of parameter $s \in (0, 1)$. There exists a relationship between s, b and ε that guarantees the expectation of the weights is identically one. The following proposition, proven in Appendix A by taking $\theta > 1/2$, shows that the resulting noisy chain can be transient for certain values of b, ϵ and θ.

Proposition 2.2. *Consider a geometric target density as in* (11) *and a proposal density as in* (12). *In addition, let the weights when $N = 1$ be given by* (13). *Then, for some b, ε and θ the chain generated by the noisy kernel $\tilde{P}_{N=1}$ is transient.*

In contrast, since the weights are uniformly bounded by b, the pseudo-marginal chain inherits geometric ergodicity for any θ, b and ϵ. The left plot in Figure 7 shows an example. We will discuss the behaviour of this example as N increases in Section 3.4 .

2.4 Transient noisy chain with non-homogeneous weights

One could argue that the transient behaviour of the previous example is related to the large value of θ in the proposal distribution. However, as shown here, for any value of $\theta \in (0, 1)$ one can construct weights satisfying (2) for which the noisy chain is transient. With the same assumptions as in the example in Section 2.3, except that now the distribution of weights is not homogeneous but given by

$$W_{m,1} = (b - \varepsilon_m)Ber(s_m) + \varepsilon_m, \quad \text{for } b > 1 \text{ and } \varepsilon_m = m^{-(3-(m \text{ (mod 3)}))}, \tag{14}$$

the noisy chain will be transient for b large enough. The proof can be found in Appendix A.

Proposition 2.3. *Consider a geometric target density as in* (11) *and a proposal density as in* (12). *In addition, let the weights when $N = 1$ be given by* (14). *Then, for any $\theta \in (0, 1)$ there exists some $b > 1$ such that the chain generated by the noisy kernel $\tilde{P}_{N=1}$ is transient.*

The reason for this becomes apparent when looking at the behaviour of the ratios of weights. Even though $\varepsilon_m \to 0$ as $m \to \infty$, the non-monotonic behaviour of the sequence implies

$$\frac{\varepsilon_{m-1}}{\varepsilon_m} \in \begin{cases} O\left(m^2\right) & \\ O\left(m^{-1}\right) & \end{cases} \quad \text{if} \quad \begin{array}{l} m \text{ (mod 3)} = 0, \\ m \text{ (mod 3)} \in \{1, 2\}, \end{array}$$

and

$$\frac{\varepsilon_{m+1}}{\varepsilon_m} \in \begin{cases} O\left(m^{-2}\right) & \\ O\left(m\right) & \end{cases} \quad \text{if} \quad \begin{array}{l} m \text{ (mod 3)} = 2, \\ m \text{ (mod 3)} \in \{0, 1\}. \end{array}$$

Hence, the ratio of the weights can become arbitrarily large or arbitrarily close to zero with a non-negligible probability. This allows the algorithm to accept moves to the right more often, if m is large enough. Once again, the pseudo-marginal chain inherits the geometrically ergodic property from the marginal. See the central and right plots of Figure 7 for two examples using different proposals. Again, we will come back to this example in Section 3.4, where we look at the behaviour of the associated noisy chain as N increases.
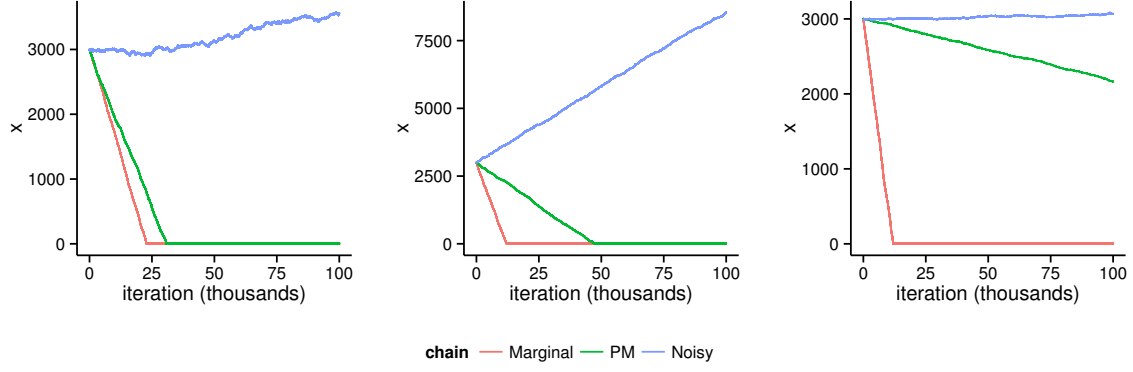
Figure 7: Runs of the marginal, pseudo-marginal and noisy chains. Left plot shows example in Section 2.3, where $\theta = 0.75$, $\varepsilon = 2 - \sqrt{3}$ and $b = 2\varepsilon\frac{\theta}{1-\theta}$. Central and right plots show example in Section 2.4, where $\theta = 0.5$ and $\theta = 0.25$ respectively, with $\varepsilon_m = m^{-(3-m \pmod 3)}$ and $b = 3 + \left(\frac{1-\theta}{\theta}\right)^3$.

3 Inheritance of ergodic properties

The inheritance of various ergodic properties of the marginal chain by pseudo-marginal Markov chains has been established using techniques that are powerful but suitable only for reversible Markov chains (see, e.g. Andrieu and Vihola, 2015). Since the noisy Markov chains treated here can be non-reversible, a suitable tool for establishing geometric ergodicity is the use of Foster–Lyapunov functions, via geometric drift towards a small set.

Definition 3.1 (Small set). Let P be the transition kernel of a Markov chain $\mathbf{\Phi}$. A subset $C \subseteq \mathcal{X}$ is small if there exists a positive integer n_0, $\varepsilon > 0$ and a probability measure $\nu(\cdot)$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that the following minorisation condition holds

$$P^{n_0}(x, \cdot) \geq \varepsilon\nu(\cdot), \qquad \text{for } x \in C. \tag{15}$$

The following theorem, which is immediate from combining Roberts and Rosenthal (1997, Proposition 2.1) and Meyn and Tweedie (2009, Theorem 15.0.1), establishes the equivalence between geometric ergodicity and a geometric drift condition. For any kernel $K : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0, 1]$, let

$$KV(x) := \int_{\mathcal{X}} K(x, dz)V(z).$$

Theorem 3.1. *Suppose that Φ is a ϕ-irreducible and aperiodic Markov chain with transition kernel P and invariant distribution π. Then, the following statements are equivalent:*

i. *There exists a small set C, constants $\lambda < 1$ and $b < \infty$, and a function $V \geq 1$ finite for some $x_0 \in \mathcal{X}$ satisfying the geometric drift condition*

$$PV(x) \leq \lambda V(x) + b\mathbb{1}_{\{x \in C\}}, \quad \text{for } x \in \mathcal{X}. \tag{16}$$

ii. *The chain is π-a.e. geometrically ergodic, meaning that for π-a.e. $x \in \mathcal{X}$ it satisfies (8) for some $V \geq 1$ (which can be taken as in (i)) and constants $\tau < 1$, $R < \infty$.*

From this point onwards, it is assumed that the marginal and noisy chains are ϕ-irreducible and aperiodic. In addition, for many of the following results, it is required that

(P1) The marginal chain is geometrically ergodic, implying its kernel P satisfies the geometric drift condition in (16) for some constants $\lambda < 1$ and $b < \infty$, some function $V \geq 1$ and a small set $C \subseteq \mathcal{X}$.

3.1 Conditions involving a negative moment

From the examples of the previous section, it is clear that the weights plays a fundamental role in the behaviour of the noisy chain. The following theorem states that the noisy chain will inherit geometric ergodicity from the marginal under some conditions on the weights involving a strengthened version of the Law of Large Numbers and convergence of negative moments.

(W1) For any $\delta > 0$, the weights $\{W_{x,N}\}_{x,N}$ satisfy

$$\lim_{N \to \infty} \sup_{x \in \mathcal{X}} \mathbb{P}_{Q_{x,N}} \left[\left| W_{x,N} - 1 \right| \geq \delta \right] = 0.$$

(W2) The weights $\{W_{x,N}\}_{x,N}$ satisfy

$$\lim_{N \to \infty} \sup_{x \in \mathcal{X}} \mathbb{E}_{Q_{x,N}} \left[W_{x,N}^{-1} \right] = 1.$$

Theorem 3.2. *Assume (P1), (W1) and (W2). Then, there exists $N_0 \in \mathbb{N}^+$ such that for all $N \geq N_0$, the noisy chain with transition kernel $\tilde{P}_N$ is geometrically ergodic.*

The above result is obtained by controlling the dissimilarity of the marginal and noisy kernels. This is done by looking at the corresponding rejection and acceptance probabilities. The proofs of the following lemmas appear in Appendix A.

Lemma 3.1. *For any $\delta > 0$*

$$\mathbb{P}_{Q_{z,N} \otimes Q_{x,N}} \left[\frac{W_{z,N}}{W_{x,N}} \leq 1 - \delta \right] \leq 2 \sup_{x \in \mathcal{X}} \mathbb{P}_{Q_{x,N}} \left[\left| W_{x,N} - 1 \right| \geq \frac{\delta}{2} \right].$$

Lemma 3.2. *Let $\rho(x)$ and $\tilde{\rho}_N(x)$ be the rejection probabilities as defined in (5) and (7) respectively. Then, for any $\delta > 0$*

$$\tilde{\rho}_N(x) - \rho(x) \leq \delta + 2 \sup_{x \in \mathcal{X}} \mathbb{P}_{Q_{x,N}} \left[\left| W_{x,N} - 1 \right| \geq \frac{\delta}{2} \right].$$

Lemma 3.3. *Let $\alpha(x,y)$ and $\tilde{\alpha}_N(x,y)$ be the acceptance probabilities as defined in (5) and (6) respectively. Then,*

$$\tilde{\alpha}_N(x,y) \leq \alpha(x,y) \mathbb{E}_{Q_{x,N}} \left[W_{x,N}^{-1} \right].$$

Notice that (W1) and (W2) allow control on the bounds in the above lemmas. While Lemma 3.2 provides a bound for the difference of the rejection probabilities, Lemma 3.3 gives one for the ratio of the acceptance probabilities. The proof of Theorem 3.2 is now presented.

Proof of Theorem 3.2. Since the marginal chain P is geometrically ergodic, it satisfies the geometric drift condition in (16) for some $\lambda < 1$, $b < \infty$, some function $V \geq 1$ and a small set $C \subseteq \mathcal{X}$. Now, using the above lemmas

$$\tilde{P}_N V(x) - PV(x) = \int_{\mathcal{X}} q(x, dz) \left(\tilde{\alpha}_N(x, z) - \alpha(x, z) \right) V(z) + V(x) \left(\tilde{\rho}_N(x) - \rho(x) \right)$$

$$\leq \left(\sup_{x \in \mathcal{X}} \mathbb{E} \left[W_{x,N}^{-1} \right] - 1 \right) PV(x) + \left(\delta + 2 \sup_{x \in \mathcal{X}} \mathbb{P} \left[\left| W_{x,N} - 1 \right| \geq \frac{\delta}{2} \right] \right) V(x).$$

By (W1) and (W2), for any $\varepsilon, \delta > 0$ there exists $N_0 \in \mathbb{N}^+$ such that

$$\sup_{x \in \mathcal{X}} \mathbb{P} \left[\left| W_{x,N} - 1 \right| \geq \frac{\delta}{2} \right] < \frac{\varepsilon}{4} \quad \text{and} \quad \sup_{x \in \mathcal{X}} \mathbb{E} \left[W_{x,N}^{-1} \right] - 1 < \varepsilon,$$

whenever $N \geq N_0$, implying

$$\tilde{P}_N V(x) \leq PV(x) + \varepsilon PV(x) + \left(\delta + \frac{\varepsilon}{2} \right) V(x)$$

$$\leq \lambda \left(1 + \delta + \frac{\varepsilon}{2} \right) V(x) + b \left(1 + \varepsilon \right) \mathbb{1}_{\{x \in C\}}.$$

Taking $\delta = \frac{\varepsilon}{2}$ and $\varepsilon \in \left(0, \frac{1-\lambda}{\lambda} \right)$, the noisy chain $\tilde{P}_N$ also satisfies a geometric drift condition for the same function V and small set C, completing the proof. $\qquad \square$

3.2 Conditions on the proposal distribution

In this subsection a different bound for the acceptance probabilities is provided, which allows dropping assumption (W2) but imposes a different one on the proposal q instead.

> **(P1*)** (P1) holds and for the same drift function V in (P1) there exists $K < \infty$ such that the proposal kernel q satisfies
> $$qV(x) \leq KV(x), \quad \text{for } x \in \mathcal{X}.$$

Theorem 3.3. *Assume (P1*) and (W1). Then, there exists $N_0 \in \mathbb{N}^+$ such that for all $N \geq N_0$, the noisy chain with transition kernel $\tilde{P}_N$ is geometrically ergodic.*

In order to prove Theorem 3.3 the following lemma is required. Its proof can be found in Appendix A. In contrast with Lemma 3.3, this lemma provides a bound for the additive difference of the noisy and marginal acceptance probabilities.

Lemma 3.4. *Let $\alpha(x, y)$ and $\tilde{\alpha}_N(x, y)$ be the acceptance probabilities as defined in (5) and (6), respectively. Then, for any $\eta > 0$*
$$\tilde{\alpha}_N(x, y) - \alpha(x, y) \leq \eta + 2 \sup_{x \in \mathcal{X}} \mathbb{P}_{Q_{x,N}} \left[\left| W_{x,N} - 1 \right| \geq \frac{\eta}{2(1 + \eta)} \right].$$

Proof of Theorem 3.3. Using Lemma 3.2 and Lemma 3.4 with $\eta = \delta$

$$\begin{aligned}
\tilde{P}_N & V(x) - PV(x) \\
&= \int_{\mathcal{X}} q(x, dz) \left(\tilde{\alpha}_N(x, z) - \alpha(x, z) \right) V(z) + V(x) \left(\tilde{\rho}_N(x) - \rho(x) \right) \\
&\leq \left(\delta + 2 \sup_{x \in \mathcal{X}} \mathbb{P} \left[\left| W_{x,N} - 1 \right| \geq \frac{\delta}{2(1 + \delta)} \right] \right) qV(x) + \left(\delta + 2 \sup_{x \in \mathcal{X}} \mathbb{P} \left[\left| W_{x,N} - 1 \right| \geq \frac{\delta}{2} \right] \right) V(x) \\
&\leq \left(\delta + 2 \sup_{x \in \mathcal{X}} \mathbb{P} \left[\left| W_{x,N} - 1 \right| \geq \frac{\delta}{2(1 + \delta)} \right] \right) (qV(x) + V(x)).
\end{aligned}$$

By (W1), there exists $N_0 \in \mathbb{N}^+$ such that

$$\sup_{x \in \mathcal{X}} \mathbb{P} \left[\left| W_{x,N} - 1 \right| \geq \frac{\delta}{2(1 + \delta)} \right] < \frac{\varepsilon}{4},$$

whenever $N \geq N_0$. This implies

$$\tilde{P}_N V(x) \leq PV(x) + \left(\delta + \frac{\varepsilon}{2} \right) (qV(x) + V(x)),$$

and using (P1*)

$$\tilde{P}_N V(x) \leq \left(\lambda + \left(\delta + \frac{\varepsilon}{2} \right) (K + 1) \right) V(x) + b \mathbb{1}_{\{x \in C\}}.$$

Taking $\delta = \frac{\varepsilon}{2}$ and $\varepsilon \in \left(0, \frac{1 - \lambda}{1 + K} \right)$, the noisy chain $\tilde{P}_N$ also satisfies a geometric drift condition for the same function V and small set C, completing the proof. $\qquad\square$

In general, assumption (P1*) may be difficult to verify as one must identify a particular function V, but it is easily satisfied when restricting to log-Lipschitz targets and when using a random walk proposal of the form

$$q(x, dy) = q(\|y - x\|)dy, \tag{17}$$

where $\| \cdot \|$ denotes the usual Euclidean distance. To see this the following assumption is required, which is a particular case of (P1) and is satisfied under some extra technical conditions (see, e.g., Roberts and Tweedie, 1996).

(P1**) $\mathcal{X} \subseteq \mathbb{R}^d$. The target π is log-Lipschitz, meaning that for some $L > 0$

$$|\log \pi(z) - \log \pi(x)| \leq L\|z - x\|.$$

(P1) holds taking the drift function $V = \pi^{-s}$, for any $s \in (0,1)$. The proposal q is a random walk as in (17) satisfying

$$\int_{\mathbb{R}^d} \exp\{a\|u\|\} \, q(\|u\|) du < \infty,$$

for some $a > 0$.

See Appendix A for a proof of the following proposition.

Proposition 3.1. *Assume (P1**) and (W1). Then, (P1*) holds.*

3.3 Conditions for arithmetic averages

In the particular setting where the weights are given by (4), sufficient conditions on these can be obtained to ensure geometric ergodicity is inherited by the noisy chain. For the simple case where the weights are homogeneous with respect to the state space (W1) is automatically satisfied. In order to attain (W2), the existence of a negative moment for a single weight is required. See Appendix A for a proof of the following result.

Proposition 3.2. *Assume weights as in* (4). *If* $\mathbb{E}_{Q_x}\left[W_x^{-1}\right] < \infty$ *then*

$$\lim_{N \to \infty} \mathbb{E}_{Q_{x,N}}\left[W_{x,N}^{-1}\right] = 1. \tag{18}$$

For homogeneous weights, (18) implies (W2). When the weights are not homogeneous, stronger conditions are needed for (W1) and (W2) to be satisfied. An appropriate first assumption is that the weights are uniformly integrable.

(W3) The weights $\{W_x\}_x$ satisfy

$$\lim_{K \to \infty} \sup_{x \in \mathcal{X}} \mathbb{E}_{Q_x}\left[W_x \mathbb{1}_{\{W_x > K\}}\right] = 0.$$

The second condition imposes an additional assumption on the distribution of the weights $\{W_x\}_x$ near 0.

(W4) There exists $\gamma \in (0,1)$ and constants $M < \infty$, $\beta > 0$ such that for $w \in (0,\gamma)$ the weights $\{W_x\}_x$ satisfy

$$\sup_{x \in \mathcal{X}} \mathbb{P}_{Q_x}[W_x \leq w] \leq Mw^{\beta}.$$

These new conditions ensure (W1) and (W2) are satisfied.

Proposition 3.3. *For weights as in* (4),

  i. *(W3) implies (W1);*

  ii. *(W1) and (W4) imply (W2).*

The following corollary is obtained as an immediate consequence of the above proposition, Theorem 3.2 and Theorem 3.3.

Corollary 3.1. *Let the weights be as in* (4). *Assume (W3) and either*

  i. *(P1) and (W4);*

  ii. *(P1*).*

Then, there exists $N_0 \in \mathbb{N}^+$ such that for all $N \geq N_0$, the noisy chain with transition kernel $\tilde{P}_N$ is geometrically ergodic.

The proof of Proposition 3.3 follows the statement of Lemma 3.5, whose proof can be found in Appendix A. This lemma allows us to characterise the distribution of $W_{x,N}$ near 0 assuming (W4) and also provides conditions for the existence and convergence of negative moments.

Lemma 3.5. *Let $\gamma \in (0,1)$ and $p > 0$.*

  i. *Suppose Z is a positive random variable, and assume that for $z \in (0,\gamma)$*

$$\mathbb{P}\left[Z \leq z\right] \leq M z^\alpha, \quad \text{where } \alpha > p, M < \infty.$$

  Then,

$$\mathbb{E}\left[Z^{-p}\right] \leq \frac{1}{\gamma^p} + pM \frac{\gamma^{\alpha-p}}{\alpha-p}.$$

  ii. *Suppose $\{Z_i\}_{i=1}^N$ is a collection of positive and independent random variables, and assume that for each $i \in \{1, \ldots, N\}$ and $z \in (0,\gamma)$*

$$\mathbb{P}\left[Z_i \leq z\right] \leq M_i z^{\alpha_i}, \quad \text{where } \alpha_i > 0, M_i < \infty.$$

  Then, for $z \in (0,\gamma)$

$$\mathbb{P}\left[\sum_{i=1}^N Z_i \leq z\right] \leq \prod_{i=1}^N M_i z^{\sum_{i=1}^N \alpha_i}.$$

  iii. *Let the weights be as in (4). If for some $N_0 \in \mathbb{N}^+$*

$$\mathbb{E}_{Q_{x,N_0}}\left[W_{x,N_0}^{-p}\right] < \infty,$$

  then for any $N \geq N_0$

$$\mathbb{E}_{Q_{x,N+1}}\left[W_{x,N+1}^{-p}\right] \leq \mathbb{E}_{Q_{x,N}}\left[W_{x,N}^{-p}\right].$$

  iv. *Assume (W1) and let $g : \mathbb{R}^+ \to \mathbb{R}$ be a function that is continuous at 1 and bounded on the interval $[\gamma, \infty)$. Then*

$$\lim_{N \to \infty} \sup_{x \in \mathcal{X}} \mathbb{E}_{Q_{x,N}}\left[|g\left(W_{x,N}\right) - g\left(1\right)|\mathbb{1}_{W_{x,N} \geq \gamma}\right] = 0.$$

Proof of Proposition 3.3. Part (i) is a consequence of Chandra (1989, Theorem 1). Assuming (W3), it implies

$$\lim_{N \to \infty} \sup_{x \in \mathcal{X}} \mathbb{E}|W_{x,N} - 1| = 0.$$

By Markov's inequality

$$\mathbb{E}|W_{x,N} - 1| \geq \delta \mathbb{P}\left[|W_{x,N} - 1| \geq \delta\right],$$

and the result follows.

To prove (ii), assume (W4) and by part (ii) of Lemma 3.5, for $w \in (0,\gamma)$

$$\mathbb{P}\left[NW_{x,N} \leq w\right] \leq M^N w^{N\beta}.$$

Take $p > 1$ and define $N_0 := \lfloor \frac{p}{\beta} \rfloor + 1$, then using part (i) of Lemma 3.5 if $N \geq N_0$

$$\sup_{x \in \mathcal{X}} \mathbb{E}\left[W_{x,N}^{-p}\right] \leq \frac{N}{\gamma^p} + pNM^N \frac{\gamma^{N\beta-p}}{N\beta-p}.$$

Hence, by Hölder's inequality

$$\mathbb{E}\left[|W_{x,N}^{-1} - 1|\mathbb{1}_{W_{x,N} \in (0,\gamma)}\right] \leq \mathbb{E}\left[W_{x,N}^{-1}\mathbb{1}_{W_{x,N} \in (0,\gamma)}\right]$$

$$\leq \left(\mathbb{E}\left[W_{x,N}^{-p}\right]\right)^{\frac{1}{p}} \left(\mathbb{P}\left[W_{x,N} < \gamma\right]\right)^{\frac{p-1}{p}},$$

and applying part (iii) of Lemma 3.5, for $N \geq N_0$

$$\mathbb{E}\left[\left|W_{x,N}^{-1} - 1\right| \mathbb{1}_{W_{x,N} \in (0,\gamma)}\right] \leq \left(\mathbb{E}\left[W_{x,N_0}^{-p}\right]\right)^{\frac{1}{p}} \left(\mathbb{P}\left[W_{x,N} < \gamma\right]\right)^{\frac{p-1}{p}}.$$

Therefore,

$$\sup_{x \in \mathcal{X}} \mathbb{E}\left[\left|W_{x,N}^{-1} - 1\right| \mathbb{1}_{W_{x,N} \in (0,\gamma)}\right] \leq \left(\sup_{x \in \mathcal{X}} \mathbb{E}\left[W_{x,N_0}^{-p}\right]\right)^{\frac{1}{p}} \left(\sup_{x \in \mathcal{X}} \mathbb{P}\left[W_{x,N} < \gamma\right]\right)^{\frac{p-1}{p}}.$$

Since $\gamma < 1$ and by (W1)

$$\lim_{N \to \infty} \sup_{x \in \mathcal{X}} \mathbb{P}\left[W_{x,N} < \gamma\right] = 0,$$

implying

$$\lim_{N \to \infty} \sup_{x \in \mathcal{X}} \mathbb{E}\left[\left|W_{x,N}^{-1} - 1\right| \mathbb{1}_{W_{x,N} \in (0,\gamma)}\right] = 0. \tag{19}$$

Now, for fixed $\gamma \in (0,1)$ the function $g(x) = x^{-1}$ is bounded and continuous on $[\gamma, \infty)$, implying by part (iv) of Lemma 3.5

$$\lim_{N \to \infty} \sup_{x \in \mathcal{X}} \mathbb{E}\left[\left|W_{x,N}^{-1} - 1\right| \mathbb{1}_{W_{x,N} \in [\gamma,\infty)}\right] = 0. \tag{20}$$

Finally, using (19) and (20)

$$\lim_{N \to \infty} \sup_{x \in \mathcal{X}} \mathbb{E}\left|W_{x,N}^{-1} - 1\right| = 0,$$

and by the triangle inequality

$$\sup_{x \in \mathcal{X}} \mathbb{E}\left|W_{x,N}^{-1} - 1\right| \geq \sup_{x \in \mathcal{X}} \mathbb{E}\left[W_{x,N}^{-1}\right] - 1,$$

the result follows. □

3.4 Remarks on results

Equipped with these results, we return to the examples in Section 2.3 and Section 2.4. Even though the noisy chain can be transient in these examples, the behaviour is quite different when considering weights that are arithmetic averages of the form in (4). Since in both examples the weights are uniformly bounded by the constant b, they immediately satisfy (W1). Additionally, by Proposition 3.2, condition (W2) is satisfied for the example in Section 2.3. This is not the case for example in Section 2.4, but condition (P1*) is satisfied by taking $V = \pi^{-\frac{1}{2}}$. Therefore, applying Theorem 3.2 and Theorem 3.3 to examples in Section 2.3 and in Section 2.4 respectively, as N increases the corresponding chains will go from being transient to geometrically ergodic.

Despite conditions (W1) and (W2) guaranteeing the inheritance of geometric ergodicity for the noisy chain, they are not necessary. Consider a modification of the example in Section 2.3, where the weights are given by

$$W_{m,1} = (b_m - \varepsilon_m) Ber(s_m) + \varepsilon_m, \quad \text{where } b_m > 1 \text{ and } \varepsilon_m \in (0,1] \text{ for all } m \geq 1.$$

Again, there exists a relationship between the variables b_m, ε_m and s_m for ensuring the expectation of the weights is equal to one. Let $Bin(N,s)$ denote a binomial distribution of parameters $N \in \mathbb{N}^+$ and $s \in (0,1)$. Then, in the arithmetic average context, $W_{m,N}$ becomes

$$W_{m,N} = \frac{(b_m - \varepsilon_m)}{N} Bin(N, s_m) + \varepsilon_m, \quad \text{where } b_m > 1 \text{ and } \varepsilon_m \in (0,1] \text{ for all } m \geq 1. \tag{21}$$

For particular choices of the sequences $\{b_m\}_{m \in \mathbb{N}^+}$ and $\{\varepsilon_m\}_{m \in \mathbb{N}^+}$, the resulting noisy chain can be geometrically ergodic for all $N \geq 1$, even though neither (W1) nor (W2) hold.

Proposition 3.4. *Consider a geometric target density as in* (11) *and a proposal density as in* (12)*. In addition, let the weights be as in* (21) *with* $b_m \to \infty$, $\varepsilon_m \to 0$ *as* $m \to \infty$ *and*

$$\lim_{m \to \infty} \frac{\varepsilon_{m-1}}{\varepsilon_m} = l, \quad \text{where } l \in \mathbb{R}^+ \cup \{+\infty\}.$$

Then, the chain generated by the noisy kernel $\tilde{P}_N$ *is geometrically ergodic for any* $N \in \mathbb{N}^+$.

Finally, in many of the previous examples, increasing the value of N seems to improve the ergodic properties of the noisy chain. However, the geometric ergodicity property is not always inherited, no matter how large N is taken. The following proposition shows an example rather similar to Proposition 3.4, but in which the ratio $\frac{\varepsilon_{m-1}}{\varepsilon_m}$ does not converge as $m \to \infty$.

Proposition 3.5. *Consider a geometric target density as in* (11) *and a proposal density as in* (12)*. In addition, let the weights be as in* (21) *with* $b_m = m$ *and*

$$\varepsilon_m = m^{-(3-(m \pmod 3))}.$$

Then, the chain generated by the noisy kernel $\tilde{P}_N$ *is transient for any* $N \in \mathbb{N}^+$.

4 Convergence of the noisy invariant distribution

So far the only concern has been whether the noisy chain inherits the geometric ergodicity property from the marginal chain. As an immediate consequence, geometric ergodicity guarantees the existence of an invariant probability distribution $\tilde{\pi}_N$ for $\tilde{P}_N$, provided N is large enough. In addition, using the same conditions from Section 3, we can characterise and in some cases quantify the convergence in total variation of $\tilde{\pi}_N$ towards the desired target π, as $N \to \infty$.

4.1 Convergence in total variation

The following definition, taken from Roberts et al. (1998), characterises a class of kernels satisfying a geometric drift condition as in (16) for the same V, C, λ and b.

Definition 4.1 (Simultaneous geometric ergodicity)**.** A class of Markov chain kernels $\{P_k\}_{k \in \mathcal{K}}$ is simultaneously geometrically ergodic if there exists a class of probability measures $\{\nu_k\}_{k \in \mathcal{K}}$, a measurable set $C \subseteq \mathcal{X}$, a real valued measurable function $V \geq 1$, a positive integer n_0 and positive constants ε, λ, b such that for each $k \in \mathcal{K}$:

    i. C is small for P_k, with $P_k^{n_0}(x, \cdot) \geq \varepsilon \nu_k(\cdot)$ for all $x \in C$;

    ii. the chain P_k satisfies the geometric drift condition in (16) with drift function V, small set C and constants λ and b.

Provided N is large, the noisy kernels $\{\tilde{P}_{N+k}\}_{k \geq 0}$ together with the marginal P will be simultaneous geometrically ergodic. This will allow the use of coupling arguments for ensuring $\tilde{\pi}_N$ and π get arbitrarily close in total variation. The main additional assumption is

    (P2) For some $\varepsilon > 0$, some probability measure $\nu(\cdot)$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and some subset $C \subseteq \mathcal{X}$, the marginal acceptance probability α and the proposal kernel q satisfy

$$\alpha(x, y) q(x, dy) \geq \varepsilon \nu(dy), \qquad \text{for } x \in C.$$

Remark 4.1. (P2) ensures the marginal chain satisfies the minorisation condition in (15), purely attained by the sub-kernel $\alpha(x, y) q(x, dy)$. This occurs under fairly mild assumptions (see, e.g., Roberts and Tweedie, 1996, Theorem 2.2).

Theorem 4.1. *Assume (P1), (P2), (W1) and (W2). Alternatively, assume (P1*), (P2) and (W1). Then,*

    i. *there exists* $N_0 \in \mathbb{N}^+$ *such that the class of kernels* $\left\{P, \tilde{P}_{N_0}, \tilde{P}_{N_0+1}, \dots\right\}$ *is simultaneously geometrically ergodic;*

ii. for all $x \in \mathcal{X}$, $\lim_{N \to \infty} \|\tilde{P}_N(x, \cdot) - P(x, \cdot)\|_{TV} = 0$;

iii. $\lim_{N \to \infty} \|\tilde{\pi}_N(\cdot) - \pi(\cdot)\|_{TV} = 0$.

Part (iii) of the above theorem is mainly a consequence of Roberts et al. (1998, Theorem 9) when parts (i) and (ii) hold. Indeed, by the triangle inequality,

$$\|\tilde{\pi}_N(\cdot) - \pi(\cdot)\|_{TV} \leq \|\tilde{P}_N^n(x, \cdot) - \tilde{\pi}_N(\cdot)\|_{TV} + \|P^n(x, \cdot) - \pi(\cdot)\|_{TV} + \|\tilde{P}_N^n(x, \cdot) - P^n(x, \cdot)\|_{TV}. \qquad (22)$$

Provided $N \geq N_0$, the first two terms in (22) can be made arbitrarily small by increasing n. In addition, due to the simultaneous geometrically ergodic property, the first term in (22) is uniformly controlled regardless the value of N. Finally, using an inductive argument, part (ii) implies that for all $x \in \mathcal{X}$ and all $n \in \mathbb{N}^+$

$$\lim_{N \to \infty} \|\tilde{P}_N^n(x, \cdot) - P^n(x, \cdot)\|_{TV} = 0.$$

Proof of Theorem 4.1. From the proofs of Theorem 3.2 and Theorem 3.3, there exists $N_2 \in \mathbb{N}^+$ such that the class of kernels $\left\{P, \tilde{P}_{N_2}, \tilde{P}_{N_2+1}, \dots\right\}$ satisfies condition (ii) in Definition 4.1 for the same function V, small set C and constants λ_{N_2}, b_{N_2}. Respecting (i), for any $\delta \in (0, 1)$

$$\begin{aligned}
\tilde{P}_N(x, A) &\geq \int_A \tilde{\alpha}_N(x, z) q(x, dz) \\
&\geq \int_A \mathbb{E}\left[\min\left\{1, \frac{W_{z,N}}{W_{x,N}}\right\}\right] \alpha(x, z) q(x, dz) \\
&\geq (1 - \delta) \int_A \left(1 - \mathbb{P}\left[\frac{W_{z,N}}{W_{x,N}} \leq 1 - \delta\right]\right) \alpha(x, z) q(x, dz).
\end{aligned}$$

Then, by Lemma 3.1

$$\tilde{P}_N(x, A) \geq (1 - \delta)\left(1 - 2 \sup_{x \in \mathcal{X}} \mathbb{P}\left[\left|W_{x,N} - 1\right| \geq \frac{\delta}{2}\right]\right) \int_A \alpha(x, z) q(x, dz).$$

By (W1), there exists $N_1 \in \mathbb{N}^+$ such that for $N \geq N_1$

$$\sup_{x \in \mathcal{X}} \mathbb{P}\left[\left|W_{x,N} - 1\right| \geq \frac{\delta}{2}\right] \leq \frac{\delta}{2},$$

giving

$$\tilde{P}_N(x, A) \geq (1 - \delta)^2 \int_A \alpha(x, z) q(x, dz).$$

Due to (P2),

$$\tilde{P}_N(x, A) \geq (1 - \delta)^2 \varepsilon \nu(A), \qquad \text{for } x \in C.$$

Finally, take $N_0 = \max\{N_1, N_2\}$ implying (i).

To prove (ii) apply Lemma 3.2 and Lemma 3.4 to get

$$\begin{aligned}
\sup_{A \in \mathcal{B}(\mathcal{X})} \left\{\tilde{P}_N(x, A) - P(x, A)\right\} &\leq \left(\eta + 2 \sup_{x \in \mathcal{X}} \mathbb{P}_{Q_{x,N}}\left[\left|W_{x,N} - 1\right| \geq \frac{\eta}{2(1 + \eta)}\right]\right) \sup_{A \in \mathcal{B}(\mathcal{X})} q(x, A) \\
&\quad + (\tilde{\rho}_N(x) - \rho(x)) \sup_{A \in \mathcal{B}(\mathcal{X})} \mathbb{1}_{x \in A} \\
&\leq \left(\eta + 2 \sup_{x \in \mathcal{X}} \mathbb{P}_{Q_{x,N}}\left[\left|W_{x,N} - 1\right| \geq \frac{\eta}{2(1 + \eta)}\right]\right) \\
&\quad + \left(\delta + 2 \sup_{x \in \mathcal{X}} \mathbb{P}_{Q_{x,N}}\left[\left|W_{x,N} - 1\right| \geq \frac{\delta}{2}\right]\right) \qquad (23)
\end{aligned}$$

Finally, taking $N \to \infty$ and by (W1)

$$\lim_{N \to \infty} \sup_{A \in \mathcal{B}(\mathcal{X})} \left\{\tilde{P}_N(x, A) - P(x, A)\right\} \leq \eta + \delta.$$

The result follows since η and δ can be taken arbitrarily small.

For (iii), see Theorem 9 in Roberts et al. (1998) for a detailed proof. $\qquad \square$

4.2 Rate of convergence

Let $(\tilde{\Phi}_n^N)_{n\geq 0}$ denote the noisy chain and $(\Phi_n)_{n\geq 0}$ the marginal chain, which move according to the kernels $\tilde{P}_N$ and P, respectively. Let $c_x := 1 - \|\tilde{P}_N(x,\cdot) - P(x,\cdot)\|_{TV}$, using notions of maximal coupling for random variables defined on a Polish space (see Lindvall 2002 and Thorisson 2013). In particular, there exists a probability measure $\nu_x(\cdot)$ such that

$$P(x,\cdot) \geq c_x \nu_x(\cdot) \quad \text{and} \quad \tilde{P}_N(x,\cdot) \geq c_x \nu_x(\cdot).$$

Let $c := \inf_{x \in \mathcal{X}} c_x$, define a coupling in the following way

- If $\tilde{\Phi}_{n-1}^N = \Phi_{n-1} = y$, with probability c draw $\Phi_n \sim \nu_y(\cdot)$ and set $\tilde{\Phi}_n^N = \Phi_n$. Otherwise, draw independently $\Phi_n \sim R(y,\cdot)$ and $\tilde{\Phi}_n^N \sim \tilde{R}_N(y,\cdot)$, where

$$R(y,\cdot) := (1-c)^{-1}\left(P(y,\cdot) - c\nu_y(\cdot)\right) \quad \text{and}$$
$$\tilde{R}_N(y,\cdot) := (1-c)^{-1}\left(\tilde{P}_N(y,\cdot) - c\nu_y(\cdot)\right).$$

- If $\tilde{\Phi}_{n-1}^N \neq \Phi_{n-1}$, draw independently $\Phi_n \sim P(y,\cdot)$ and $\tilde{\Phi}_n^N \sim \tilde{P}_N(y,\cdot)$.

Since

$$\mathbb{P}\left[\tilde{\Phi}_n^N \neq \Phi_n | \tilde{\Phi}_0^N = \Phi_0 = x\right]$$
$$\leq \mathbb{P}\left[\tilde{\Phi}_n^N \neq \Phi_n | \tilde{\Phi}_{n-1}^N = \Phi_{n-1}, \tilde{\Phi}_0^N = \Phi_0 = x\right] + \mathbb{P}\left[\tilde{\Phi}_{n-1}^N \neq \Phi_{n-1} | \tilde{\Phi}_0^N = \Phi_0 = x\right]$$
$$\leq 1 - c + \mathbb{P}\left[\tilde{\Phi}_{n-1}^N \neq \Phi_{n-1} | \tilde{\Phi}_0^N = \Phi_0 = x\right],$$

and noting

$$\mathbb{P}\left[\tilde{\Phi}_1^N \neq \Phi_1 | \tilde{\Phi}_0^N = \Phi_0 = x\right] \leq \sup_{x \in \mathcal{X}} \|\tilde{P}_N(x,\cdot) - P(x,\cdot)\|_{TV}$$
$$= 1 - c,$$

an induction argument can be applied to obtain

$$\mathbb{P}\left[\tilde{\Phi}_n^N \neq \Phi_n | \tilde{\Phi}_0^N = \Phi_0 = x\right] \leq n \sup_{x \in \mathcal{X}} \|\tilde{P}_N(x,\cdot) - P(x,\cdot)\|_{TV}.$$

Therefore, using the coupling inequality, the third term in (22) can be bounded by

$$\|\tilde{P}_N^n(x,\cdot) - P^n(x,\cdot)\|_{TV} \leq \mathbb{P}\left[\tilde{\Phi}_n^N \neq \Phi_n | \tilde{\Phi}_0^N = \Phi_0 = x\right]$$
$$\leq n \sup_{x \in \mathcal{X}} \|\tilde{P}_N(x,\cdot) - P(x,\cdot)\|_{TV}. \tag{24}$$

On the other hand, using the simultaneous geometric ergodicity of the kernels and provided N is large enough, the noisy and marginal kernels will each satisfy a geometric drift condition as in (16) with a common drift function $V \geq 1$, small set C and constants λ, b. Therefore, by Theorem 3.1, there exist $R > 0$, and $\tau < 1$ such that

$$\|\tilde{P}_N^n(x,\cdot) - \tilde{\pi}_N(\cdot)\|_{TV} \leq RV(x)\tau^n \quad \text{and} \quad \|P^n(x,\cdot) - \pi(\cdot)\|_{TV} \leq RV(x)\tau^n. \tag{25}$$

Explicit values for R and τ are in principle possible, as done in Rosenthal (1995) and Meyn and Tweedie (1994). For simplicity assume $\inf_{x \in \mathcal{X}} V(x) = 1$, then combining (24) and (25) in (22), for all $n \in \mathbb{N}^+$

$$\|\tilde{\pi}_N(\cdot) - \pi(\cdot)\|_{TV} \leq 2R\tau^n + n \sup_{x \in \mathcal{X}} \|\tilde{P}_N(x,\cdot) - P(x,\cdot)\|_{TV}. \tag{26}$$

So, if an analytic expression in terms of N is available for the second term on the right hand side of (26), it will be possible to obtain an explicit rate of convergence for $\tilde{\pi}_N$ and π.

Theorem 4.2. *Assume (P1), (P2), (W1) and (W2). Alternatively, assume (P1*), (P2) and (W1). In addition, suppose*

$$\sup_{x \in \mathcal{X}} \|\tilde{P}_N(x, \cdot) - P(x, \cdot)\|_{TV} \leq \frac{1}{r(N)},$$

where $r : \mathbb{N}^+ \to \mathbb{R}^+$ and $\lim_{N \to \infty} r(N) = +\infty$. Then, there exists $D > 0$ and $N_0 \in \mathbb{N}^+$ such that for all $N \geq N_0$,

$$\|\tilde{\pi}_N(\cdot) - \pi(\cdot)\|_{TV} \leq D \frac{\log(r(N))}{r(N)}.$$

Proof. Let $R > 0, \tau \in (0, 1)$ and $r > 0$. Pick r large enough, such that

$$\log\left(2Rr\log\left(\tau^{-1}\right)\right) \geq 1,$$

then the convex function $f : [1, \infty) \to \mathbb{R}^+$ where

$$f(s) = 2R\tau^s + \frac{s}{r},$$

is minimised at

$$s_* = \frac{\log\left(2Rr\log\left(\tau^{-1}\right)\right)}{\log\left(\tau^{-1}\right)}.$$

Restricting the domain of f to the positive integers and due to convexity, it is then minimised at either

$$n_1 = \lfloor s_* \rfloor \qquad \text{or} \qquad n_2 = \lceil s_* \rceil.$$

In any case

$$\min\{f(n_1), f(n_2)\} \leq f(s_* + 1),$$

$$= \frac{1}{r}\left(1 + \frac{\tau + \log\left(2Rr\log\left(\tau^{-1}\right)\right)}{\log\left(\tau^{-1}\right)}\right).$$

Finally take N large enough such that

$$\log\left(2Rr(N)\log\left(\tau^{-1}\right)\right) \geq 1,$$

and from (26)

$$\|\tilde{\pi}_N(\cdot) - \pi(\cdot)\|_{TV} \leq \min\{f(n_1), f(n_2)\}$$

$$\leq \frac{1}{r(N)}\left(1 + \frac{\tau + \log\left(2Rr(N)\log\left(\tau^{-1}\right)\right)}{\log\left(\tau^{-1}\right)}\right)$$

$$= O\left(\frac{\log(r(N))}{r(N)}\right),$$

obtaining the result. $\qquad \square$

Moreover, when the weights are expressed in terms of arithmetic averages as in (4), an explicit expression for $r(N)$ can be obtained whenever there exists a uniformly bounded moment. This is a slightly stronger assumption than (W3).

(W5) There exists $k > 0$, such that the weights $\{W_x\}_x$ satisfy

$$\sup_{x \in \mathcal{X}} \mathbb{E}_{Q_x}\left[W_x^{1+k}\right] < \infty.$$

Proposition 4.1. *Assume (P1), (P2), (W4) and (W5). Alternatively, assume (P1*), (P2) and (W5). Then, there exists $D_k > 0$ and $N_0 \in \mathbb{N}^+$ such that for all $N \geq N_0$,*

$$\|\tilde{\pi}_N(\cdot) - \pi(\cdot)\|_{TV} \leq D_k \frac{\log(N)}{N^{1 - \frac{2}{2+k}}}.$$

If in addition (W5) holds for all $k > 0$, then for any $\varepsilon \in (0, 1)$ there will exist $D_\varepsilon > 0$ and $N_0 \in \mathbb{N}^+$ such that for all $N \geq N_0$,

$$\|\tilde{\pi}_N(\cdot) - \pi(\cdot)\|_{TV} \leq D_\varepsilon \frac{\log(N)}{N^{1-\varepsilon}}.$$

5 Discussion

In this article, fundamental stability properties of the noisy algorithm have been explored. The noisy Markov kernels considered are perturbed Metropolis–Hastings kernels defined by a collection of state-dependent distributions for non-negative weights all with expectation 1. The general results do not assume a specific form for these weights, which can be simple arithmetic averages or more complex random variables. The former may arise when unbiased importance sampling estimates of a target density are used, while the latter may arise when such densities are estimated unbiasedly using a particle filter.

Two different sets of sufficient conditions were provided under which the noisy chain inherits geometric ergodicity from the marginal chain. The first pair of conditions, (W1) and (W2), involve a stronger version of the Law of Large Numbers for the weights and uniform convergence of the first negative moment, respectively. For the second set, (W1) is still required but (W2) can be replaced with (P1*), which imposes a condition on the proposal distribution. These conditions also imply simultaneous geometric ergodicity of a sequence of noisy Markov kernels together with the marginal Markov kernel, which then ensures that the noisy invariant $\tilde{\pi}_N$ converges to π in total variation as N increases. Moreover, an explicit bound for the rate of convergence between $\tilde{\pi}_N$ and π is possible whenever an explicit bound (that is uniform in x) is available for the convergence between $\tilde{P}_N(x, \cdot)$ and $P(x, \cdot)$.

When weights are arithmetic averages as in (4), specific conditions were given for inheriting geometric ergodicity from the corresponding marginal chain. The uniform integrability condition in (W3) ensures that (W1) is satisfied, whereas (W4) is essential for satisfying (W2). Regarding the noisy invariant distribution $\tilde{\pi}_N$, (W5), which is slightly stronger than (W3), leads to an explicit bound on the rate of convergence of this distribution to π.

The noisy algorithm remains undefined when the weights have positive probability of being zero. If both weights were zero one could accept the move, reject the move or keep sampling new weights until one of them is not zero. Each of these lead to different behaviour.

As seen in the examples of Section 3.4, the behaviour of the ratio of the weights (at least in the tails of the target) plays an important role in the ergodic properties of the noisy chain. In this context, it seems plausible to obtain geometric noisy chains, even when the marginal is not, if the ratio of the weights decays sufficiently fast to zero in the tails. Another interesting possibility, that may lead to future research, is to relax the condition on the expectation of the weights to be identically one.

Acknowledgements

References

P. Alquier, N. Friel, R. Everitt, and A. Boland. Noisy Monte Carlo: convergence of Markov chains with approximate transition kernels. *Statistics and Computing*, pages 1–19, 2014. ISSN 0960-3174. doi: 10.1007/s11222-014-9521-x. URL http://dx.doi.org/10.1007/s11222-014-9521-x.

C. Andrieu and G. O. Roberts. The Pseudo-Marginal Approach for Efficient Monte Carlo Computations. *The Annals of Statistics*, 37(2):pp. 697–725, 2009. ISSN 00905364. URL http://www.jstor.org/stable/30243645.

C. Andrieu and M. Vihola. Establishing some order amongst exact approximations of MCMCs. *arXiv preprint arXiv:1404.6909*, 2014.

C. Andrieu and M. Vihola. Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Probab.*, 25(2):1030–1077, 04 2015. doi: 10.1214/14-AAP1022. URL http://dx.doi.org/10.1214/14-AAP1022.

C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 72(3):pp. 269–342, 2010. ISSN 13697412. URL http://www.jstor.org/stable/40802151.

M. A. Beaumont. Estimation of Population Growth or Decline in Genetically Monitored Populations. *Genetics*, 164(3):1139–1160, 2003. URL http://www.genetics.org/content/164/3/1139.abstract.

H. Callaert and J. Keilson. On exponential ergodicity and spectral structure for birth-death processes, II. *Stochastic Processes and their Applications*, 1(3):217 – 235, 1973. ISSN 0304-4149. doi: http://dx.doi.org/10.1016/0304-4149(73)90001-X. URL http://www.sciencedirect.com/science/article/pii/030441497390001X.

K. S. Chan and C. J. Geyer. Discussion: Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, 22(4):pp. 1747–1758, 1994. ISSN 00905364. URL http://www.jstor.org/stable/2242481.

T. K. Chandra. Uniform Integrability in the Cesàro Sense and the Weak Law of Large Numbers. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 51(3):pp. 309–317, 1989. ISSN 0581572X. URL http://www.jstor.org/stable/25050754.

P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Probability and Its Applications. Springer, 2004. ISBN 9780387202686. URL http://books.google.co.uk/books?id=8LypfuG8ZLYC.

J. Fernández-Villaverde and J. F. Rubio-Ramírez. Estimating Macroeconomic Models: A Likelihood Approach. *The Review of Economic Studies*, 74(4):1059–1087, 2007. doi: 10.1111/j.1467-937X.2007.00437.x. URL http://restud.oxfordjournals.org/content/74/4/1059.abstract.

J. M. Flegal and G. L. Jones. Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38(2):pp. 1034–1070, 2010. ISSN 00905364. URL http://www.jstor.org/stable/25662268.

M. Girolami, A.-M. Lyne, H. Strathmann, D. Simpson, and Y. Atchade. Playing Russian roulette with intractable likelihoods. *arXiv preprint arXiv:1306.4032*, 2013.

A. Khuri and G. Casella. The Existence of the First Negative Moment Revisited. *The American Statistician*, 56(1):pp. 44–47, 2002. ISSN 00031305. URL http://www.jstor.org/stable/3087326.

A. Lee and K. Łatuszyński. Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation. *Biometrika*, 2014. doi: 10.1093/biomet/asu027. URL http://biomet.oxfordjournals.org/content/early/2014/08/05/biomet.asu027.abstract.

T. Lindvall. *Lectures on the Coupling Method*. Dover Books on Mathematics Series. Dover Publications, Incorporated, 2002. ISBN 9780486421452. URL http://books.google.co.uk/books?id=GUwyU1ypd1wC.

F. Maire, R. Douc, and J. Olsson. Comparison of asymptotic variances of inhomogeneous Markov chains with application to Markov chain Monte Carlo methods. *Ann. Statist.*, 42(4):1483–1510, 08 2014. doi: 10.1214/14-AOS1209. URL http://dx.doi.org/10.1214/14-AOS1209.

T. J. McKinley, J. V. Ross, R. Deardon, and A. R. Cook. Simulation-based Bayesian inference for epidemic models. *Computational Statistics & Data Analysis*, 71(0):434 – 447, 2014. ISSN 0167-9473. doi: http://dx.doi.org/10.1016/j.csda.2012.12.012. URL http://www.sciencedirect.com/science/article/pii/S016794731200446X.

K. L. Mengersen and R. L. Tweedie. Rates of Convergence of the Hastings and Metropolis Algorithms. *The Annals of Statistics*, 24(1):pp. 101–121, 1996. ISSN 00905364. URL http://www.jstor.org/stable/2242610.

S. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009. ISBN 0521731828, 9780521731829.

S. P. Meyn and R. L. Tweedie. Computable Bounds for Geometric Convergence Rates of Markov Chains. *The Annals of Applied Probability*, 4(4):pp. 981–1011, 1994. ISSN 10505164. URL http://www.jstor.org/stable/2245077.

J. Norris. *Markov Chains*. Number no. 2008 in Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, 1999. ISBN 9780521633963. URL https://books.google.co.uk/books?id=qM65VRmOJZAC.

P. D. O'Neill, D. J. Balding, N. G. Becker, M. Eerola, and D. Mollison. Analyses of Infectious Disease Data from Household Outbreaks by Markov chain Monte Carlo Methods. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 49(4):pp. 517–542, 2000. ISSN 00359254. URL http://www.jstor.org/stable/2680786.

W. W. Piegorsch and G. Casella. The Existence of the First Negative Moment. *The American Statistician*, 39(1):pp. 60–62, 1985. ISSN 00031305. URL http://www.jstor.org/stable/2683910.

M. K. Pitt, R. dos Santos Silva, P. Giordani, and R. Kohn. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, 171(2):134 – 151, 2012. ISSN 0304-4076. doi: http://dx.doi.org/10.1016/j.jeconom.2012.06.004. URL http://www.sciencedirect.com/science/article/pii/S0304407612001510. Bayesian Models, Methods and Applications.

G. Roberts and J. Rosenthal. Geometric Ergodicity and Hybrid Markov Chains. *Electron. Commun. Probab.*, 2:no. 2, 13–25, 1997. ISSN 1083-589X. doi: 10.1214/ECP.v2-981. URL http://ecp.ejpecp.org/article/view/981.

G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.

G. O. Roberts and R. L. Tweedie. Geometric Convergence and Central Limit Theorems for Multidimensional Hastings and Metropolis Algorithms. *Biometrika*, 83(1):pp. 95–110, 1996. ISSN 00063444. URL http://www.jstor.org/stable/2337435.

G. O. Roberts, J. S. Rosenthal, and P. O. Schwartz. Convergence Properties of Perturbed Markov Chains. *Journal of Applied Probability*, 35(1):pp. 1–11, 1998. ISSN 00219002. URL http://www.jstor.org/stable/3215541.

J. S. Rosenthal. Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo. *Journal of the American Statistical Association*, 90(430):pp. 558–566, 1995. ISSN 01621459. URL http://www.jstor.org/stable/2291067.

C. Sherlock, A. H. Thiery, G. O. Roberts, and J. S. Rosenthal. On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.*, 43(1):238–275, 02 2015. doi: 10.1214/14-AOS1278. URL http://dx.doi.org/10.1214/14-AOS1278.

H. Thorisson. *Coupling, Stationarity, and Regeneration*. Probability and Its Applications. Springer New York, 2013. ISBN 9781461270553. URL http://books.google.co.uk/books?id=187hnQEACAAJ.

A Proofs

A.1 On state-dependent random walks

The following proposition for state-dependent Markov chains on the positive integers will be useful for addressing some proofs. See Norris (1999) for a proof of parts (*i*) and (*ii*), for part (*iii*) see Callaert and Keilson (1973), which is proved within the birth-death process context.

Proposition A.1. *Suppose we have a random walk Φ on $\mathbb{N}^+$ with transition kernel P. Define for $m \geq 1$*

$$p_m := P(m, \{m+1\}) \quad and \quad q_m := P(m, \{m-1\}),$$

with $q_1 = 0, p_1 \in (0,1]$ and $p_m, q_m > 0, p_m + q_m \leq 1$ for all $m \geq 2$. The resulting chain is:

i. recurrent if and only if

$$\sum_{m=2}^{\infty} \prod_{i=2}^{m} \frac{q_i}{p_i} \to \infty;$$

ii. positive recurrent if and only if

$$\sum_{m=2}^{\infty} \prod_{i=2}^{m} \frac{p_{i-1}}{q_i} < \infty;$$

iii. geometrically ergodic if

$$\lim_{m \to \infty} p_m < \lim_{m \to \infty} q_m.$$

Remark A.1. Notice that (3) is not an if and only if statement and that it implies (2). Additionally, if the chain is not state-dependent, (2) implies (3).

A.2 Section 2

Proof of Proposition 2.1. Since h is convex

$$h(m) - h(m-1) \geq h'(m-1) \quad \text{and} \quad h(m) - h(m+1) \leq -h'(m),$$

implying

$$\frac{\tilde{P}_N(m, \{m-1\})}{\tilde{P}_N(m, \{m+1\})} \geq \frac{\mathbb{E}\left[\min\left\{1, \exp\{h'(m-1)\}\frac{W_N^{(1)}}{W_N^{(2)}}\right\}\right]}{\mathbb{E}\left[\min\left\{1, \exp\{-h'(m)\}\frac{W_N^{(1)}}{W_N^{(2)}}\right\}\right]}.$$

Define $Z := \frac{W_N^{(1)}}{W_N^{(2)}}$, and since $\pi(m) \to 0$ it is true that

$$\log(k) := \lim_{m \to \infty} h'(m) > 0, \tag{27}$$

hence

$$\lim_{m \to \infty} \frac{\tilde{P}_N(m, \{m-1\})}{\tilde{P}_N(m, \{m+1\})} \geq \frac{\mathbb{E}\left[\min\{1, kZ\}\right]}{\mathbb{E}\left[\min\{1, k^{-1}Z\}\right]}. \tag{28}$$

If $k = +\infty$, it is clear that the limit in (28) diverges, consequently the noisy chain is geometrically ergodic according to Proposition A.1. If $k < \infty$, the noisy chain will be geometrically ergodic if

$$\mathbb{E}\left[\min\{1, kZ\}\right] > \mathbb{E}\left[\min\{1, k^{-1}Z\}\right],$$

which can be translated to

$$k\mathbb{E}\left[Z\mathbb{1}_{\{Z \leq k^{-1}\}}\right] + \mathbb{P}\left[Z > k^{-1}\right] \qquad > k^{-1}\mathbb{E}\left[Z\mathbb{1}_{\{Z < k\}}\right] + \mathbb{P}\left[Z \geq k\right],$$

or equivalently to

$$k\mathbb{P}\left[k^{-1} < Z < k\right] + \left(k^2 - 1\right)\mathbb{E}\left[Z\mathbb{1}_{\{Z \leq k^{-1}\}}\right] > \mathbb{E}\left[Z\mathbb{1}_{\{k^{-1} < Z < k\}}\right]. \tag{29}$$

Now consider two cases, first if $\mathbb{P}\left[k^{-1} < Z < k\right] > 0$ then it is clear that

$$\mathbb{E}\left[(k - Z)\mathbb{1}_{\{k^{-1} < Z < k\}}\right] > 0,$$

which satisfies (29). Finally, if $\mathbb{P}\left[k^{-1} < Z < k\right] = 0$ then

$$\mathbb{P}\left[Z \leq k^{-1}\right] = \frac{1}{2} = \mathbb{P}\left[Z \geq k\right],$$

implying from (27)

$$\left(k^2 - 1\right)\mathbb{E}\left[Z\mathbb{1}_{\{Z \leq k^{-1}\}}\right] > 0,$$

and leading to (29). □

Proof of Proposition 2.2. For simplicity the subscript N is dropped. In this case,

$$Q_{m,1} = Q = \mathcal{L}\left((b - \varepsilon)Ber(s) + \varepsilon\right),$$

and the condition $\mathbb{E}_Q[W] = 1$ implies

$$s = \frac{1 - \varepsilon}{b - \varepsilon}. \tag{30}$$

Let $\theta \in \left(\frac{1}{1+2\varepsilon}, 1\right)$ and set

$$b = \varepsilon \frac{2\theta}{1 - \theta}, \tag{31}$$

this implies $\bar{\alpha}(m, w; m - 1, u) \equiv 1$ and

$$\bar{\alpha}(m, w; m + 1, u) = \begin{cases} \frac{1-\theta}{2\theta} & u = w \\ 1 & \text{if} \quad u = b, w = \varepsilon \\ \left(\frac{1-\theta}{2\theta}\right)^2 & u = \varepsilon, w = b \end{cases}.$$

Therefore, for $m \geq 2$,

$$\tilde{\alpha}(m, m - 1) = 1 \quad \text{and}$$

$$\tilde{\alpha}(m, m + 1) = \frac{1 - \theta}{2\theta} \left(s^2 + (1 - s)^2\right) + \left(1 + \left(\frac{1 - \theta}{2\theta}\right)^2\right) s(1 - s).$$

Consequently, $\tilde{P}(m, \{m - 1\}) = 1 - \theta$ and

$$\tilde{P}(m, \{m + 1\}) = \theta \left(\frac{1 - \theta}{2\theta} \left(s^2 + (1 - s)^2\right) + \left(1 + \left(\frac{1 - \theta}{2\theta}\right)^2\right) s(1 - s)\right)$$

$$> \theta s(1 - s).$$

From Proposition A.1, if

$$\tilde{P}(m, \{m + 1\}) > \tilde{P}(m, \{m - 1\}),$$

then the noisy chain will be transient. For this to happen, it is enough to pick θ and s such that

$$\theta s(1 - s) - (1 - \theta) \geq 0.$$

Let $s = \varepsilon$, then from (30) and (31)

$$\theta = \frac{(1 - \varepsilon + \varepsilon^2)}{1 - \varepsilon + 3\varepsilon^2} \tag{32}$$

$$= 1 - \frac{2\varepsilon^2}{1 - \varepsilon + 3\varepsilon^2},$$

and if $\varepsilon \leq 2 - \sqrt{3}$ then

$$\theta s(1 - s) - (1 - \theta) = \frac{\varepsilon}{1 - \varepsilon + 3\varepsilon^2} \left((1 - \varepsilon + \varepsilon^2)(1 - \varepsilon) - 2\varepsilon\right)$$

$$\geq \frac{\varepsilon}{1 - \varepsilon + 3\varepsilon^2} \left((1 - \varepsilon)^2 - 2\varepsilon\right)$$

$$= \frac{\varepsilon}{1 - \varepsilon + 3\varepsilon^2} \left((2 - \varepsilon)^2 - 3\right)$$

$$\geq 0.$$

Hence, for $\varepsilon \in \left(0, 2 - \sqrt{3}\right)$ and setting $s = \varepsilon$, θ as in (32) and b as in (31), the resulting noisy chain is transient. $\square$

Proof of Proposition 2.3. For simplicity the subscript N is dropped. In this case,

$$Q_{m,1} = Q_m = \mathcal{L}\left((b - \varepsilon_m)Ber(s_m) + \varepsilon_m\right),$$

and the condition $\mathbb{E}_{Q_m}[W_m] = 1$ implies

$$s_m = \frac{1 - \varepsilon_m}{b - \varepsilon_m}.$$

Then, for m large enough

$$
\begin{aligned}
\tilde{\alpha}(m, m-1) &= \mathbb{E}\left[\bar{\alpha}(m, W_m; m-1, W_{m-1})\right] \\
&= \min\left\{1, \frac{2\theta}{1-\theta}\right\} s_{m-1}s_m + s_{m-1}(1 - s_m) + (1 - s_{m-1})(1 - s_m)\mathbb{1}_{\{m \pmod 3 = 0\}} \\
&\quad + O\left(m^{-1}\right) \quad \text{and} \\
\tilde{\alpha}(m, m-1) &= \mathbb{E}\left[\bar{\alpha}(m, W_m; m-1, W_{m-1})\right] \\
&= \min\left\{1, \frac{1-\theta}{2\theta}\right\} s_m s_{m+1} + (1 - s_m)s_{m+1} + (1 - s_m)(1 - s_{m+1})\mathbb{1}_{\{m \pmod 3 \neq 2\}} \\
&\quad + O\left(m^{-1}\right).
\end{aligned}
$$

Define

$$
\begin{aligned}
c_m &:= \frac{\tilde{P}(m, \{m-1\})}{\tilde{P}(m, \{m+1\})} \\
&= \frac{(1-\theta)\tilde{\alpha}(m, m-1)}{\theta\tilde{\alpha}(m, m+1)}.
\end{aligned}
$$

Since $s_m \to \frac{1}{b}$ as $m \to \infty$,

$$
\begin{aligned}
c_{0,\infty} &:= \lim_{k \to \infty} c_{3k} \\
&= \left(\frac{1-\theta}{\theta}\right) \frac{\left(\min\left\{1, \frac{2\theta}{1-\theta}\right\} - 1\right)\frac{1}{b^2} + \frac{1}{b} + \left(1 - \frac{1}{b}\right)^2}{\left(\min\left\{1, \frac{1-\theta}{2\theta}\right\} - 1\right)\frac{1}{b^2} + \frac{1}{b} + \left(1 - \frac{1}{b}\right)^2} \\
&\leq \left(\frac{1-\theta}{\theta}\right) \frac{1}{1 - \frac{1}{b}} \\
&= \left(\frac{1-\theta}{\theta}\right) \frac{b}{b-1} =: l_0,
\end{aligned}
$$

$$
\begin{aligned}
c_{1,\infty} &:= \lim_{k \to \infty} c_{3k}c_{3k+1} \\
&= c_{0,\infty}\left(\frac{1-\theta}{\theta}\right) \frac{\left(\min\left\{1, \frac{2\theta}{1-\theta}\right\} - 1\right)\frac{1}{b^2} + \frac{1}{b}}{\left(\min\left\{1, \frac{1-\theta}{2\theta}\right\} - 1\right)\frac{1}{b^2} + \frac{1}{b} + \left(1 - \frac{1}{b}\right)^2} \\
&\leq l_0\left(\frac{1-\theta}{\theta}\right) \frac{\frac{1}{b}}{1 - \frac{1}{b}} \\
&= \left(\frac{1-\theta}{\theta}\right)^2 \frac{b}{(b-1)^2} =: l_1
\end{aligned}
$$

and

$$
\begin{aligned}
\lim_{k \to \infty} c_{3k}c_{3k+1}c_{3k+2} &= c_{1,\infty}\left(\frac{1-\theta}{\theta}\right) \frac{\left(\min\left\{1, \frac{2\theta}{1-\theta}\right\} - 1\right)\frac{1}{b^2} + \frac{1}{b}}{\left(\min\left\{1, \frac{1-\theta}{2\theta}\right\} - 1\right)\frac{1}{b^2} + \frac{1}{b}} \\
&\leq l_1\left(\frac{1-\theta}{\theta}\right) \frac{\frac{1}{b}}{\frac{1}{b}\left(1 - \frac{1}{b}\right)} \\
&= \left(\frac{1-\theta}{\theta}\right)^3 \frac{b^2}{(b-1)^3} =: l_2.
\end{aligned}
$$

Therefore, for any $\delta > 0$ there exists $k_0 \in \mathbb{N}^+$, such that whenever $k \geq k_0 + 1$

$$K := \prod_{j=k_0}^{k-1} c_{3j} c_{3j+1} c_{3j+2}$$
$$< (l_2 + \delta)^{k-k_0},$$

implying

$$Kc_{3k} < (l_2 + \delta)^{k-k_0}(l_0 + \delta),$$
$$Kc_{3k}c_{3k+1} < (l_2 + \delta)^{k-k_0}(l_1 + \delta) \quad \text{and}$$
$$Kc_{3k}c_{3k+1}c_{3k+2} < (l_2 + \delta)^{k-k_0}(l_2 + \delta).$$

Hence, for $i \in \{0, 1, 2\}$ and some $C > 0$

$$\prod_{j=2}^{3k+i} c_j < C(l_2 + \delta)^k.$$

Let $a_m := \prod_{j=2}^{m} c_j$, then a sufficient condition for the series $\sum_{m=2}^{\infty} a_m$ to converge, implying a transient chain according to Proposition A.1, is $l_2 < 1$. This is the case for $b \geq 3 + \left(\frac{1-\theta}{\theta}\right)^3$, since

$$1 - l_2 = 1 - \left(\frac{1-\theta}{\theta}\right)^3 \frac{b^2}{(b-1)^3}$$
$$= \frac{b^2}{(b-1)^3} \left(\frac{(b-1)^3}{b^2} - \left(\frac{1-\theta}{\theta}\right)^3\right)$$
$$= \frac{b^2}{(b-1)^3} \left(b - 3 + \frac{3}{b} - \frac{1}{b^2} - \left(\frac{1-\theta}{\theta}\right)^3\right)$$
$$> \frac{b^2}{(b-1)^3} \left(b - 3 - \left(\frac{1-\theta}{\theta}\right)^3\right)$$
$$\geq 0.$$

Hence, the resulting noisy chain is transient if $b \geq 3 + \left(\frac{1-\theta}{\theta}\right)^3$, for any $\theta \in (0, 1)$. $\qquad\square$

A.3   Section 3

Proof of Lemma 3.1.   For any $\delta > 0$

$$\mathbb{P}\left[\frac{W_{z,N}}{W_{x,N}} \leq 1 - \delta\right] \leq \mathbb{P}\left[W_{x,N} \geq 1 + \frac{\delta}{2}\right] + \mathbb{P}\left[W_{z,N} \leq 1 - \frac{\delta}{2}\right]$$
$$\leq \mathbb{P}\left[\left|W_{x,N} - 1\right| \geq \frac{\delta}{2}\right] + \mathbb{P}\left[\left|W_{z,N} - 1\right| \geq \frac{\delta}{2}\right]$$
$$\leq 2 \sup_{x \in \mathcal{X}} \mathbb{P}_{Q_{x,N}}\left[\left|W_{x,N} - 1\right| \geq \frac{\delta}{2}\right].$$

$\qquad\square$

Proof of Lemma 3.2.   Using the inequality

$$\min\{1, ab\} \geq \min\{1, a\}\min\{1, b\}, \quad \text{for } a, b \geq 0,$$

and applying Markov's inequality with $\delta > 0$,

$$
\begin{aligned}
\tilde{\rho}_N(x) &= 1 - \int_{\mathcal{X}} q(x, dz) \tilde{\alpha}_N(x, z) \\
&\le 1 - \int_{\mathcal{X}} q(x, dz) \alpha(x, z) \mathbb{E}\left[\min\left\{1, \frac{W_{z,N}}{W_{x,N}}\right\}\right] \\
&\le 1 - (1 - \delta) \int_{\mathcal{X}} q(x, dz) \alpha(x, z) \mathbb{P}\left[\min\left\{1, \frac{W_{z,N}}{W_{x,N}}\right\} > 1 - \delta\right] \\
&= 1 - (1 - \delta) \int_{\mathcal{X}} q(x, dz) \alpha(x, z) + (1 - \delta) \int_{\mathcal{X}} q(x, dz) \alpha(x, z) \mathbb{P}\left[\frac{W_{z,N}}{W_{x,N}} \le 1 - \delta\right] \\
&\le 1 - (1 - \delta)(1 - \rho(x)) + \int_{\mathcal{X}} q(x, dz) \alpha(x, z) \mathbb{P}\left[\frac{W_{z,N}}{W_{x,N}} \le 1 - \delta\right].
\end{aligned}
$$

Finally, using Lemma 3.1

$$
\begin{aligned}
\tilde{\rho}_N(x) &\le \rho(x) + \delta(1 - \rho(x)) + 2 \sup_{x \in \mathcal{X}} \mathbb{P}\left[\left|W_{x,N} - 1\right| \ge \frac{\delta}{2}\right](1 - \rho(x)) \\
&\le \rho(x) + \delta + 2 \sup_{x \in \mathcal{X}} \mathbb{P}\left[\left|W_{x,N} - 1\right| \ge \frac{\delta}{2}\right].
\end{aligned}
$$

$\square$

Proof of Lemma 3.3. For the first claim apply Jensen's inequality and the fact that

$$
\min\{1, ab\} \le \min\{1, a\} b, \quad \text{for } a \ge 0 \text{ and } b \ge 1,
$$

hence

$$
\begin{aligned}
\tilde{\alpha}_N(x, z) &\le \min\left\{1, \frac{\pi(z) q(z, x)}{\pi(x) q(x, z)} \mathbb{E}\left[\frac{W_{z,N}}{W_{x,N}}\right]\right\} \\
&\le \alpha(x, z) \mathbb{E}\left[W_{x,N}^{-1}\right] \mathbb{E}\left[W_{z,N}\right].
\end{aligned}
$$

$\square$

Proof of Lemma 3.4. Using the inequality

$$
\min\{1, ab\} \le \min\{1, a\} b, \quad \text{for } a \ge 0 \text{ and } b \ge 1,
$$

$$
\begin{aligned}
\tilde{\alpha}_N(x, z) &= \mathbb{E}\left[\bar{\alpha}_N(x, W_{x,N}; z, W_{z,N}) \mathbb{1}_{\left\{\frac{W_{z,N}}{W_{x,N}} < 1 + \eta\right\}}\right] + \mathbb{E}\left[\bar{\alpha}_N(x, W_{x,N}; z, W_{z,N}) \mathbb{1}_{\left\{\frac{W_{z,N}}{W_{x,N}} \ge 1 + \eta\right\}}\right] \\
&\le \alpha(x, z)(1 + \eta) \mathbb{P}\left[\frac{W_{z,N}}{W_{x,N}} < 1 + \eta\right] + \mathbb{P}\left[\frac{W_{z,N}}{W_{x,N}} \ge 1 + \eta\right] \\
&\le \alpha(x, z) + \eta + \mathbb{P}\left[\frac{W_{z,N}}{W_{x,N}} \ge 1 + \eta\right],
\end{aligned}
$$

Notice that

$$
\mathbb{P}\left[\frac{W_{z,N}}{W_{x,N}} \ge 1 + \eta\right] = \mathbb{P}\left[\frac{W_{x,N}}{W_{z,N}} \le \frac{1}{1 + \eta}\right],
$$

then applying Lemma 3.1 taking $\delta = \frac{\eta}{1+\eta}$.

$$
\tilde{\alpha}_N(x, z) \le \alpha(x, z) + \eta + 2 \sup_{x \in \mathcal{X}} \mathbb{P}\left[\left|W_{x,N} - 1\right| \ge \frac{\eta}{2(1 + \eta)}\right].
$$

$\square$

Proof of Proposition 3.1. Taking $V = \pi^{-s}$, where $0 < s < \min\left\{1, \frac{a}{L}\right\}$,

$$\frac{qV(x)}{V(x)} = \int_{\mathcal{X}} \frac{V(z)}{V(x)} q(x, dz)$$

$$= \int_{\mathcal{X}} \left(\frac{\pi(x)}{\pi(z)}\right)^s q(x, dz)$$

$$\leq \int_{\mathbb{R}^d} \exp\{a\|z - x\|\} q(\|z - x\|) dz.$$

Finally, using the transformation $u = z - x$,

$$\frac{qV(x)}{V(x)} \leq \int_{\mathbb{R}^d} \exp\{a\|u\|\} q(\|u\|) du,$$

which implies (P1*). $\qquad\square$

Proof of Proposition 3.2. By properties of the arithmetic and harmonic means

$$\frac{1}{N} \sum_{k=1}^{N} \frac{1}{W_x^{(k)}} - \frac{N}{\sum_{k=1}^{N} W_x^{(k)}} \geq 0,$$

which implies, by Jensen's inequality,

$$\mathbb{E}\left[\frac{1}{N} \sum_{k=1}^{N} \frac{1}{W_x^{(k)}} - \frac{N}{\sum_{k=1}^{N} W_x^{(k)}}\right] \leq \mathbb{E}\left[W_x^{-1}\right] - 1.$$

Then, using Fatou's lemma and the law of large numbers

$$\mathbb{E}\left[W_x^{-1}\right] - 1 \geq \limsup_{N\to\infty} \mathbb{E}\left[\frac{1}{N} \sum_{k=1}^{N} \frac{1}{W_x^{(k)}} - W_{x,N}^{-1}\right]$$

$$\geq \liminf_{N\to\infty} \mathbb{E}\left[\frac{1}{N} \sum_{k=1}^{N} \frac{1}{W_x^{(k)}} - W_{x,N}^{-1}\right]$$

$$\geq \mathbb{E}\left[\liminf_{N\to\infty}\left(\frac{1}{N} \sum_{k=1}^{N} \frac{1}{W_x^{(k)}}\right) - \limsup_{N\to\infty} W_{x,N}^{-1}\right]$$

$$\geq \mathbb{E}\left[W_x^{-1}\right] - 1,$$

hence

$$\lim_{N\to\infty} \mathbb{E}\left[\frac{1}{N} \sum_{k=1}^{N} \frac{1}{W_x^{(k)}} - W_{x,N}^{-1}\right] = E\left[W_x^{-1}\right] - 1. \tag{33}$$

Finally, since

$$\mathbb{E}\left[\frac{1}{N} \sum_{k=1}^{N} \frac{1}{W_x^{(k)}} - W_{x,N}^{-1}\right] = E\left[W_x^{-1}\right] - E\left[W_{x,N}^{-1}\right],$$

the expression in (33) becomes

$$\lim_{N\to\infty} \mathbb{E}\left[W_{x,N}^{-1}\right] = 1.$$

$\qquad\square$

Proof of Lemma 3.5. The proof of (i) is motivated by Piegorsch and Casella (1985, Theorem 2.1) and Khuri and Casella (2002, Theorem 3), however the existence of a density function is not assumed here.

Since Z^{-p} is positive,

$$
\begin{aligned}
\mathbb{E}\left[Z^{-p}\right] &= \int_{\mathbb{R}^+} \mathbb{P}\left[Z^{-p} \geq z\right] dz \\
&\leq \frac{1}{\gamma^p} + \int_{(\gamma^{-p},\infty)} \mathbb{P}\left[Z^{-p} \geq z\right] dz \\
&= \frac{1}{\gamma^p} + \int_{(0,\gamma)} pu^{-p-1}\mathbb{P}\left[Z \leq u\right] du \\
&\leq \frac{1}{\gamma^p} + pM\frac{\gamma^{\alpha-p}}{\alpha - p}.
\end{aligned}
$$

For part (ii), since the random variables $\{Z_i\}$ are positive, then for any $z > 0$

$$
\mathbb{P}\left[\sum_{i=1}^N Z_i \leq z\right] = \mathbb{P}\left[\sum_{i=1}^N Z_i \leq z, \max_{i\in\{1,\ldots,N\}}\{Z_i\} \leq z\right].
$$

Therefore, for $z \in (0,\gamma)$

$$
\begin{aligned}
\mathbb{P}\left[\sum_{i=1}^N Z_i \leq z\right] &\leq \mathbb{P}\left[\max_{i\in\{1,\ldots,N\}}\{Z_i\} \leq z\right] \\
&= \prod_{i=1}^N \mathbb{P}\left[Z_i \leq z\right] \\
&\leq \prod_{i=1}^N M_i z^{\sum_{i=1}^N \alpha_i}.
\end{aligned}
$$

Part (iii) can be seen as a consequence of $W_{x,N}$ and $W_{x,N+1}$ being convex ordered and $g(x) = x^{-p}$ being a convex function for $x > 0$ and $p \geq 0$, (see, e.g., Andrieu and Vihola, 2014). We provide a self-contained proof by defining for $j \in \{1,\ldots,N+1\}$

$$
S_{x,N}^{(j)} := \frac{1}{N} \sum_{k=1,k\neq j}^{N+1} W_x^{(k)},
$$

and we have

$$
W_{x,N+1} = \frac{1}{N+1} \sum_{j=1}^{N+1} S_{x,N}^{(j)}
$$

and since the arithmetic mean is greater than or equal to the geometric mean

$$
W_{x,N+1} \geq \left(\prod_{j=1}^{N+1} S_{x,N}^{(j)}\right)^{\frac{1}{N+1}}.
$$

This implies for $p > 0$

$$
\begin{aligned}
\mathbb{E}\left[W_{x,N+1}^{-p}\right] &\leq \mathbb{E}\left[\left(\prod_{j=1}^{N+1} S_{x,N}^{(j)}\right)^{-\frac{p}{N+1}}\right] \\
&\leq \prod_{j=1}^{N+1}\left(\mathbb{E}\left[\left(S_{x,N}^{(j)}\right)^{-p}\right]\right)^{\frac{1}{N+1}} \\
&= \mathbb{E}\left[\left(S_{x,N}^{(1)}\right)^{-p}\right] \\
&= \mathbb{E}\left[W_{x,N}^{-p}\right],
\end{aligned}
$$

where Hölder's inequality has been used and the fact that the random variables $\left\{S_{x,N}^{(j)} : j \in 1, \ldots, N+1\right\}$ are identically distributed according to $Q_{x,N}$.

For part *(iv)*, let $M_\gamma = \sup_{y \in [\gamma, \infty)} |g(y)|$ and due to continuity at $y = 1$, for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\mathbb{E}\left[\left|g(W_{x,N}) - g(1)\right| \mathbb{1}_{W_{x,N} \in [\gamma, \infty)}\right] \leq 2M_\gamma \mathbb{P}\left[\gamma \leq W_{x,N} \leq 1 - \delta\right] + 2M_\gamma \mathbb{P}\left[1 + \delta \leq W_{x,N}\right]$$
$$+ \mathbb{E}\left[\left|g(W_{x,N}) - g(1)\right| \mathbb{1}_{W_{x,N} \in (1-\delta, 1+\delta)}\right]$$
$$\leq 2M_\gamma \mathbb{P}\left[\left|W_{x,N} - 1\right| \geq \delta\right] + \varepsilon \mathbb{P}\left[\left|W_{x,N} - 1\right| < \delta\right].$$

Therefore, for fixed ε and by (W1)

$$\lim_{N \to \infty} \sup_{x \in \mathcal{X}} \mathbb{E}\left[\left|g(W_{x,N}) - g(1)\right| \mathbb{1}_{W_{x,N} \in [\gamma, \infty)}\right] \leq 2M_\gamma \lim_{N \to \infty} \sup_{x \in \mathcal{X}} \mathbb{P}\left[\left|W_{x,N} - 1\right| \geq \delta\right] + \varepsilon$$
$$\leq \varepsilon,$$

obtaining the result since ε can be picked arbitrarily small. $\qquad\square$

Proof of Proposition 3.4. First notice that if $l < \infty$ then $l \geq 1$. To see this, define

$$a_m := \frac{\varepsilon_{m-1}}{\varepsilon_m},$$

then for fixed $\delta > 0$, there exists $M \in \mathbb{N}$ such that for $m \geq M$

$$a_m < l + \delta.$$

Then, for $m \geq M$

$$\frac{\varepsilon_1}{\varepsilon_m} = \prod_{j=2}^{m} a_j$$
$$< (l + \delta)^{m-M} \frac{\varepsilon_1}{\varepsilon_M},$$

and because $\varepsilon_m \to 0$, it is clear that $(l + \delta)^m \to \infty$ as $m \to \infty$. Therefore, $l + \delta > 1$ and since δ can be taken arbitrarily small, it is true that $l \geq 1$.

Now, for weights as in (21) and using a simple random walk proposal, the noisy acceptance probability can be expressed as

$$\tilde{\alpha}_N(m, m-1) = \sum_{j=0}^{N} \sum_{k=0}^{N} \min\left\{1, \frac{2\theta}{1-\theta} \frac{b_{m-1}j + (N-j)\varepsilon_{m-1}}{b_m k + (N-k)\varepsilon_m}\right\} \binom{N}{j}$$
$$\times \binom{N}{k} (s_{m-1})^j (s_m)^k (1 - s_{m-1})^{N-j} (1 - s_m)^{N-k} \tag{34}$$

and

$$\tilde{\alpha}_N(m, m+1) = \sum_{j=0}^{N} \sum_{k=0}^{N} \min\left\{1, \frac{1-\theta}{2\theta} \frac{b_{m+1}j + (N-j)\varepsilon_{m+1}}{b_m k + (N-k)\varepsilon_m}\right\} \binom{N}{j}$$
$$\times \binom{N}{k} (s_{m+1})^j (s_m)^k (1 - s_{m+1})^{N-j} (1 - s_m)^{N-k}. \tag{35}$$

Since $b_m \to \infty$, then $s_m \to 0$ as $m \to \infty$; therefore, any term in (34) and (35), for which $j + k \neq 0$, tends to zero as $m \to \infty$. Hence,

$$\tilde{\alpha}_N(m, m-1) = \min\left\{1, \frac{2\theta}{1-\theta} \times \frac{\varepsilon_{m-1}}{\varepsilon_m}\right\} (1 - s_{m-1})^N (1 - s_m)^N + o(1)$$

and

$$\tilde{\alpha}_N(m, m+1) = \min\left\{1, \frac{1-\theta}{2\theta} \times \frac{\varepsilon_{m+1}}{\varepsilon_m}\right\} (1 - s_{m+1})^N (1 - s_m)^N + o(1),$$

implying,

$$\lim_{m \to \infty} \frac{\tilde{P}_N\left(m, \{m-1\}\right)}{\tilde{P}_N\left(m, \{m+1\}\right)} = \frac{(1-\theta) \lim_{m\to\infty} \min\left\{1, \frac{2\theta}{1-\theta} \times \frac{\varepsilon_{m-1}}{\varepsilon_m}\right\}}{\theta \lim_{m\to\infty} \min\left\{1, \frac{1-\theta}{2\theta} \times \frac{\varepsilon_{m+1}}{\varepsilon_m}\right\}}. \tag{36}$$

If $l = +\infty$, (36) tends to $+\infty$, whereas if $l < \infty$

$$\lim_{m \to \infty} \frac{\tilde{P}_N\left(m, \{m-1\}\right)}{\tilde{P}_N\left(m, \{m+1\}\right)} = 2l \frac{\min\{1-\theta, 2\theta l\}}{\min\{2\theta l, 1-\theta\}}$$
$$\geq 2.$$

In any case, this implies

$$\lim_{m \to \infty} \tilde{P}_N\left(m, \{m-1\}\right) \geq 2 \lim_{m \to \infty} \tilde{P}_N\left(m, \{m+1\}\right),$$

and since

$$\lim_{m \to \infty} \tilde{P}_N\left(m, \{m-1\}\right) = \min\{1-\theta, 2\theta l\}$$
$$> 0,$$

the noisy chain is geometrically ergodic according to Proposition Proposition A.1. $\qquad\square$

Proof of Proposition 3.5. Noting that

$$\frac{\varepsilon_{m-1}}{\varepsilon_m} \in \begin{cases} O\left(m^2\right) & \\ O\left(m^{-1}\right) & \end{cases} \text{if} \quad \begin{matrix} m \ (\mathrm{mod}\ 3) = 0, \\ m \ (\mathrm{mod}\ 3) \in \{1, 2\}, \end{matrix}$$

$$\text{and} \quad \frac{\varepsilon_{m+1}}{\varepsilon_m} \in \begin{cases} O\left(m^{-2}\right) & \\ O\left(m\right) & \end{cases} \text{if} \quad \begin{matrix} m \ (\mathrm{mod}\ 3) = 2, \\ m \ (\mathrm{mod}\ 3) \in \{0, 1\}, \end{matrix}$$

expressions in (34) and (35) become

$$\tilde{\alpha}_N(m, m-1) = (1-s_{m-1})^N (1-s_m)^N \mathbb{1}_{\{m \ (\mathrm{mod}\ 3)=0\}} + O(m^{-1})$$

and

$$\tilde{\alpha}_N(m, m+1) = (1-s_{m+1})^N (1-s_m)^N \mathbb{1}_{\{m \ (\mathrm{mod}\ 3)=0,1\}} + O\left(m^{-1}\right).$$

Therefore,

$$\frac{\tilde{P}_N\left(m, \{m-1\}\right)}{\tilde{P}_N\left(m, \{m+1\}\right)} = \left(\frac{1-\theta}{\theta}\right) \frac{(1-s_{m-1})^N + O(m^{-1})}{(1-s_{m+1})^N + O(m^{-1})} \mathbb{1}_{\{m \ (\mathrm{mod}\ 3)=0\}}$$
$$+ O\left(m^{-1}\right) \mathbb{1}_{\{m \ (\mathrm{mod}\ 3)=1\}}$$
$$+ O(1) \mathbb{1}_{\{m \ (\mathrm{mod}\ 3)=2\}},$$

implying there exists $C \in \mathbb{R}^+$ such that for $j = 0, 2$

$$\lim_{k \to \infty} \frac{\tilde{P}_N\left(3k+j, \{3k+j-1\}\right)}{\tilde{P}_N\left(3k+j, \{3k+j+1\}\right)} \leq C,$$

and

$$\lim_{k \to \infty} \frac{\tilde{P}_N\left(3k+1, \{3k\}\right)}{\tilde{P}_N\left(3k+1, \{3k+2\}\right)} = 0.$$

Then, for fixed $\delta > 0$ there exists $k_0 \in \mathbb{N}^+$ such that whenever $k \geq k_0$

$$\frac{\tilde{P}_N\left(3k+j, \{3k+j-1\}\right)}{\tilde{P}_N\left(3k+j, \{3k+j+1\}\right)} < C + \delta, \quad \text{for } j = 0, 2$$

and

$$\frac{\tilde{P}_N\left(3k+1,\{3k\}\right)}{\tilde{P}_N\left(3k+1,\{3k+2\}\right)} < \delta.$$

Let

$$c_m := \frac{\tilde{P}_N(m,\{m-1\})}{\tilde{P}_N(m,\{m+1\})},$$

then for $k \geq k_0 + 1$

$$\prod_{j=2}^{3k+1} c_j = \prod_{j=1}^{k} c_{3j-1}c_{3j}c_{3j+1}$$

$$\leq \left((C+\delta)^2\,\delta\right)^{k-k_0} \prod_{j=1}^{k_0} c_{3j-1}c_{3j}c_{3j+1}.$$

Take δ small enough, such that $(C+\delta)^2\,\delta < 1$, hence

$$\sum_{k=1}^{\infty}\prod_{j=2}^{3k+1} c_j = \sum_{k=1}^{k_0}\prod_{j=2}^{3k+1} c_j + \sum_{k=k_0}^{\infty}\prod_{j=2}^{3k+1} c_j$$

$$\leq \sum_{k=1}^{k_0}\prod_{j=2}^{3k+1} c_j + \prod_{j=1}^{k_0} c_{3j-1}c_{3j}c_{3j+1} \sum_{k=k_0}^{\infty}\left((C+\delta)^2\,\delta\right)^{k-k_0}$$

$$= \sum_{k=1}^{k_0}\prod_{j=2}^{3k+1} c_j + \frac{\prod_{j=1}^{k_0} c_{3j-1}c_{3j}c_{3j+1}}{1-(C+\delta)^2\,\delta}$$

$$< \infty.$$

Similarly, it can be proved that

$$\sum_{k=0}^{\infty}\prod_{j=2}^{3k+2} c_j < \infty \quad \text{and} \quad \sum_{k=1}^{\infty}\prod_{j=2}^{3k} c_j < \infty,$$

thus

$$\sum_{m=2}^{\infty}\prod_{j=2}^{m} c_j < \infty,$$

implying the noisy chain is transient according to Proposition A.1. $\qquad\square$

A.4 Section 4

Proof of Proposition 4.1. From (23) and taking $\delta < \frac{1}{2}$, $\eta = \frac{\delta}{1-\delta}$

$$\sup_{x\in\mathcal{X}}\|\tilde{P}_N(x,\cdot) - P(x,\cdot)\|_{TV} \leq 3\delta + 4\sup_{x\in\mathcal{X}}\mathbb{P}\left[\left|W_{x,N}-1\right| \geq \frac{\delta}{2}\right].$$

Using Markov's inequality

$$\sup_{x\in\mathcal{X}}\|\tilde{P}_N(x,\cdot) - P(x,\cdot)\|_{TV} \leq 3\delta + 4\sup_{x\in\mathcal{X}}\mathbb{P}\left[\left|W_{x,N}-1\right|^{1+k} \geq \left(\frac{\delta}{2}\right)^{1+k}\right]$$

$$\leq 3\delta + \frac{2^{3+k}}{\delta^{1+k}}\sup_{x\in\mathcal{X}}\mathbb{E}\left[\left|W_{x,N}-1\right|^{1+k}\right]$$

$$\leq 3\delta + \frac{2^{3+k}}{\delta^{1+k}N^k}\sup_{x\in\mathcal{X}}\mathbb{E}\left[\left|W_x-1\right|^{1+k}\right].$$

Now, let

$$C_k = \sup_{x \in \mathcal{X}} \mathbb{E}\left[\left|W_x - 1\right|^{1+k}\right],$$

then the convex function $f : \mathbb{R}^+ \to \mathbb{R}^+$ where

$$f(s) = 3s + \frac{2^{3+k}C_k}{s^{1+k}N^k},$$

is minimised at

$$s_* = \left(\frac{(1+k)2^{3+k}C_k}{3N^k}\right)^{\frac{1}{2+k}}$$
$$= O\left(N^{-\frac{k}{2+k}}\right).$$

Then,

$$\sup_{x \in \mathcal{X}} \|\tilde{P}_N(x, \cdot) - P(x, \cdot)\|_{TV} \leq f(s_*)$$
$$= O\left(N^{-\frac{k}{2+k}}\right) + O\left(N^{-k+\frac{k(1+k)}{2+k}}\right)$$
$$= O\left(N^{-\frac{k}{2+k}}\right).$$

Applying Theorem 4.2 by taking

$$r(N) \propto N^{1-\frac{2}{2+k}}$$

and noting $\log\left(N^{1-\frac{2}{2+k}}\right) \leq \log(N)$, the result is obtained.

For the second claim, for a given $\varepsilon \in (0,1)$ take $k_\varepsilon \geq 2\left(\varepsilon^{-1} - 1\right)$ and apply the first part. $\square$